

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 31, 2003, 02:25:34 : Search time 3121.29 Seconds
(without alignments)
4531.450 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHNRVRRVQQPVTATDSPRT.....IEEGGTASPEIPFRPMRS 486

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-O=/cgn2.1/USPTO/US09825414/runat_28012003.161741_26085/app-query.fasta.1.1294
-DB=genembl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09825414.ecgn.1.1.4707-erunat.28012003.161741_26085 -MCPU=6 -ICPU=3
-NO_XLPUT -NO_MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_pi:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	2412	100.0	52498	1 AF232004	AF232004 Pseudomon
2	1653	68.5	1834	1 AF458051	AF458051 Pseudomon
3	508	21.1	208050	1 AL646083	AL646083 Ralstonia
c 4	168	7.0	15311	1 AE007164	AE007164 Mycobacte
c 5	164	6.8	47852	1 MTY023	MTY023 Mycobacte
c 6	162.5	6.7	14131	1 AE005780	AE005780 Caulobact
7	160.5	6.7	42526	1 SC4A7	AL133423 Streptomy
8	156	6.5	10275	1 AE012164	AE012164 Xanthomon
9	155.5	6.4	13431	1 AE000234	AE000234 Escherich
10	155.5	6.4	15424	1 AE007019	AE007019 Mycobacte
11	155.5	6.4	16902	1 D90775	D90775 E.coli geno
12	155.5	6.4	18700	1 D90774	D90774 E.coli geno
13	151.5	6.3	5329	1 AF317649	AF317649 Rhodobact
14	151.5	6.3	20350	1 AL646078	AL646078 Ralstonia
15	151	6.3	40790	1 MTCV493	MTCV493 Mycobacteri
16	150.5	6.2	6413	1 AF067776	AF067776 Abiotroph
c 17	149.5	6.2	34807	1 AP003000	AP003000 Mesorhizo
18	148.5	6.2	20701	1 AE006958	AE006958 Mycobacte
19	148.5	6.2	29550	1 MTY039	MTY039 Mycobacte
20	148	6.1	12412	3 AC006669	AC006669 Caenorhab
21	148	6.1	346897	1 AP002995	AP002995 Mesorhizo
22	147.5	6.1	6744	1 SSEPFA	X71880 S.suis (189
23	147.5	6.1	6744	6 A24024	A24024 S.suis EF+
24	147.5	6.1	6744	6 I57310	I57310 Sequence 2
25	147.5	6.1	10052	1 AE004597	AE004597 Pseudomon
26	147.5	6.1	38404	1 SC2G5	AL035478 Streptomy
27	147	6.1	53662	1 MTY016	AL021841 Mycobacte
28	147	6.1	78210	1 AB070949	AB070949 Streptomy
29	147	6.1	188050	1 AL646072	AL646072 Ralstonia
30	146	6.1	15806	1 AE007151	AE007151 Mycobacte
31	145.5	6.0	10432	1 AE005333	AE005333 Escherich
32	145.5	6.0	18660	1 SC2E2	AL049645 Streptomy
33	145.5	6.0	146500	2 AC119500	AC119500 Leishmani
34	145.5	6.0	22805	1 AP002555	AP002555 Escherich
c 35	145	6.0	3701	1 AF172724	AF172724 Caulobact
c 36	145	6.0	11283	1 AE005710	AE005710 Caulobact
c 37	145	6.0	12070	1 AB011413	AB011413 Streptomy
38	145	6.0	110000	2 LMF1CHR18_07	Contribution (8 of
39	144	6.0	8126	1 AFANCC	L31363 Alcaaligenes
40	144	6.0	189050	1 AL646077	AL646077 Ralstonia
41	143.5	5.9	6761	8 AB071164	AB071164 Saccharom
42	143.5	5.9	11830	1 BPEFHAB1	M60351 B.pertussis
43	143.5	5.9	11831	1 BPEFHABH	X52156 Bordetella
44	143.5	5.9	11883	6 I06606	I06606 Sequence 1
45	143.5	5.9	46819	6 AR204176	AR204176 Sequence

RESULT 1

ALIGNMENTS

AF232004/c 52498 bp DNA linear BCT 05-MAR-2001
 LOCUS AF232004
 DEFINITION Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity island, complete sequence.
 ACCESSION AF232004 AF061028 AF061029 AF232006 L41861
 VERSION AF232004.3 GI:13325077
 KEYWORDS
 SOURCE Pseudomonas syringae pv. tomato.
 ORGANISM Pseudomonas syringae pv. tomato
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 25494 to 29778)
 AUTHORS Preston, G., Huang, H.C., He, S.Y. and Collmer, A.
 TITLE The Hrp proteins of Pseudomonas syringae pvs. syringae, glycinea, and tomato are encoded by an operon containing yerlinia ysc homologs and elicit the hypersensitive response in tomato but not soybean
 JOURNAL Mol. Plant Microbe Interact. 8 (5), 717-732 (1995)
 MEDLINE 96025089
 PUBMED 7579616
 REFERENCE 2 (bases 22134 to 25847; 29687 to 32670)
 AUTHORS Deng, W.L., Preston, G., Collmer, A., Chang, C.J. and Huang, H.C.
 TITLE Characterization of the hrpC and hrpRS operons of Pseudomonas syringae pathovars syringae, tomato, and glycinea and analysis of the ability of hrpF, hrpG, hrpC, hrpR, and hrpY mutants to elicit the hypersensitive response and disease in plants
 JOURNAL J. Bacteriol. 180 (17), 4523-4531 (1998)
 MEDLINE 98389667
 PUBMED 9721291
 REFERENCE 3 (bases 31672 to 51723)
 AUTHORS Charkowski, A.O., Alfano, J.R., Preston, G., Yuan, J., He, S.Y. and Collmer, A.
 TITLE The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and pectate lyases and can elicit the plant hypersensitive response and bind to pectate
 JOURNAL J. Bacteriol. 180 (19), 5211-5217 (1998)
 MEDLINE 98422476
 PUBMED 9748456
 REFERENCE 4 (bases 901 to 22404; 31672 to 51723)
 AUTHORS Alfano, J.R., Charkowski, A.O., Deng, W.L., Bader, J.L., Penicki-Ocwieja, T., van Dijk, K. and Collmer, A.
 TITLE The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)
 MEDLINE 20243785
 PUBMED 10781092
 REFERENCE 5 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Pseudomonas syringae pv. tomato DC3000 hrpL through hrpC unpublished
 JOURNAL 6 (bases 1 to 52498)
 REFERENCE 6 (bases 1 to 52498)
 AUTHORS Alfano, J.R. and Collmer, A.
 TITLE Direct Submission
 JOURNAL Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland Parkway, Las Vegas, NV 89154, USA
 REFERENCE 7 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334 Plant Sciences Bldg., Ithaca, NY 14850, USA
 REFERENCE 8 (bases 1 to 52498)
 AUTHORS Ramos, A.R., Rehm, A.H. and Collmer, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334 Plant Sciences Bldg., Ithaca, NY 14850, USA
 REMARK On or before Mar 14, 2001 this sequence version replaced
 COMMENT gi:3228544, gi:790906, gi:3228541, gi:8037790, gi:11276506.
 FEATURES
 SOURCE Location/Qualifiers
 1..52498

4/25/00 I have too new.

/organism="Pseudomonas syringae pv. tomato"
 /strain="DC3000"
 /db_xref="taxon:323"
 /note="exchangeable effector locus flanking the hrp/hrp cluster
 pathovar: tomato"
 complement(908..1972)
 /gene="queA"
 complement(908..1972)
 /gene="queA"
 /codon_start=1
 /transl_table=1
 /product="QueA"
 /protein_id="AAE71483.1"
 /db_xref="GI:8037772"
 /translation="MRYADTFEELPDSLTARHPLAERRSRLLTLDDPGALAHROFT
 DLRLHSGDLAMFNTRVLPARLFGKASGGKLEITVERYLDSHRILAYARASKSK
 PGSSILIDGGGEAMVARHDALEFLRAEEVLPDLVGHMPPYIDRDEGADRR
 VOYVYAOBAGVAAAPTAGLHFDQPLMEALAKGVETAFVTLHGAGTFOVREQID
 HHMSSEMLEVSODVDVAACRARGVIAVGTSVSLSEAAADCOLKPFSGDTDF
 IYGRPHVVDALVTNPHLPSTLMLVSAFAGYPTMAVMAAIEHGYRFFSYGDAM
 FTRNPAPTAQESAPEDHA"
 2090..2177
 /product="tRNA-Leu"
 2343..2753
 /note="ORF4"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="AAE71484.1"
 /db_xref="GI:8037773"
 /translation="MNKIVYKAYFKPGEVSVKVPTEGIKKGFPGDKIMKETOM
 QOTGMSQCIDGERLSKDVEDAVQNLNADGVEIQTVPLISGAVDYALKRYEIRHNR
 TETSPGOSVYEGYSGFTGVTLVAKKFSAS"
 complement(2908..3324)
 /note="ORF3"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="AAE71485.1"
 /db_xref="GI:8037774"
 /translation="MRAVKNITAKIGFPLALITIGTSLAPAVNCDLNDNSTGAT
 CGNDKDLNDVNDAAFGNDKMDNDHTDAFGNDLDDNDHTDAFGNDKND
 LDNDNKTDAFGNDRLDNDNDNNGNPPSAKK"
 complement(3939..4664)
 /note="ORF2"
 /codon_start=1
 /transl_table=1
 /product="unknown"
 /protein_id="AAE71486.1"
 /db_xref="GI:8037775"
 /translation="MOKTTLMAALAFAMLAGGVSGPAPGSDIOGAQEMKTPYKLNLD
 AYTSKRTDAVLEARTNSYNNKGLIDLVGAGFLGTPRSNMLVGSANVEPOLVDR
 GLDCAFLIDVYEAFRRSTSOQDPYRNIVQYRYKGDVDFLNKRHFETPDNAYGRVPA
 DDITAIQISPGAVSVKRLNERAKGVYLPGLVVERSMYIIPSLVDSQVSHLRTD
 YIGIYTPASRAGCTRRFLYRDG"
 5571..6410
 /gene="tnpA"
 5571..6410
 /gene="tnpA"
 /note="truncated at the amino terminus"
 /codon_start=1
 /transl_table=1
 /product="tnpA"
 /protein_id="AAE71487.1"
 /db_xref="GI:8037776"
 /translation="WGLRPAVPLAMALRIHLMQNFEDYSDPAMEALYETTLRQFA
 GLSLDRIADETTLIFRRLLKEHLAIGIIVNGYLGDRGLMIRQGVVDAITIAHP
 SSTKNDKGRDPEMHOTKRGNOYFEGKAIHGVDAESGLVSLVGTANVADYVOVO
 LHSSEETVSGDGYTGVDKRAEHODOMQSIIAARSRRKKGESLIRAYRKIEF
 TKADLRKAVEPLRVYIKRQGYTVRRGLAKNTAQOATLPSLSNLMVYKRLAMGE
 VRL"
 complement(6656..8056)

```
/note="ORF1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAJ71488.1"
/db_xref="GI:803777"
/db_xref="GI:8037778"
/translation="MRPVGPGAPGVYPTPEAEPTTAQAGNDRARSSQASSPASV
APETPLDLKRFPGARYPDMKVENIRLKEQEGRGDGVYHTRRRPDAGSSHHV
GGQSVATSSASQALQDTNFKASDLAEIMKCPSPYALAPSKAGKSQSLANV
VSILQEKHAEQRLAQGLADYVISEGDHHTINILNLEMDSCIGTSKGLAPD
SNDKLLAKAAITFDENAQKLPELAPLTKMSKDSLGMRELLDADGLVGECHNS
TSSKRELINNMKSLKASGVTLFMEHLCAESDKALNLYLSPKSGSPMARLKNYLA
OSQOQAEDELITKYNFTLVLAAGLRYVSLDPTSTYAPAEKIKRAADAMYLA
AEKIRLSPEGMVAIFGATHTATSCDGPGLAEHLGVSLVIDDGLKSRATVDINK
NYGKLPNDVRSYK"
complement(8131..10473)
/gene="hrpk"
/complement(8131..10473)
/gene="hrpk"
/codon_start=1
/transl_table=11
/product="Hrpk"
/protein_id="AAJ71489.1"
/db_xref="GI:8037778"
/translation="MRISSEPVIVNOPTEBELALAVESPLKALPTPVGGGQAGVO
FGPAGNPGAPGAEOTASSILLALLOGSGAANANNVNOVOPCADSPAEAPPT
EATAPAEVYAAPTNTDADTASNAKPADAFIDNERYSSPEALKREPVYAHLP
EREQAKELNPJAAMWARDGPDAGAMPINANPLKTAIVAAOGGNADEKIT
NKDLTFKMMKADSDADKLANYTKNPDADPQSLPKNVSAAYMRANMPLAIAADP
HHAVDAADKTVKDVNDAEDLGLASNPGLSGALQSCOSQGFQGVDEACMSG
RKAHSPGDKMFDANLEMIKKSAPTINGOPASMLSDSATLNAVAGIDISLDEKVF
DKPXSAGQAKAAVMVKLOOTQOSVIAQSLANPTKTEGGLNDRIISOLOADDVQAVL
NKSITEGRNLYRSDASLOKAVEOTKVNNSQALQTMKDKAKYNNKHPADYSGA
ISGLAQLOKDLDPDSKVPPTNQVLEKPPLOKRTASTVTNSEGAKLQGLSDK
KSDASQSLQTDNQAIYDSVLPDFTENQHNYSYTLQDSKKGKLLQGLSDE
GGPSMLTQLAEGVQKALNSIVGFASVSLIANGDKLGAQTALIDSTRMGEALIKG
IDAGKMMGREAAGLGRGCMIGRAVGMVAGEATGLAALGAIPVYVMAIDGA
LAVFGSIITIDAYVKKHAKOKAFDHNVPVLDQFGIAAH"
complement(10560..11114)
/gene="hrpl"
/complement(10560..11114)
/gene="hrpl"
/note="alternative sigma factor"
/codon_start=1
/transl_table=11
/product="Hrpl"
/protein_id="AAG3875.2"
/db_xref="GI:13325078"
/translation="MFQKTVILDSTQRPQSSSAGIRQMTADQIOMLRATIKRYMNP
DDVDILQCVLEALRNHKKFOHASKPQTLWLGIALNLIRNFRKMYRQPYQESMED
VHSELEGHGDSHOVEGHRQLARVQALDCLPSNMOKYLEVSLBMDGNTQETANSLGV
PIGTVRSRLSGARVOLKQIDPFA"
11358..12464
/gene="hrpv"
/gene="hrpv"
11358..12464
CDS
gene
CDS
Alignment Scores:
Pred. No.: 1 97e-122 Length: 52498
Score: 2412.00 Matches: 466
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-825-414-7 (1-486) x AF232004 (1-52498)
QY 1 MethIleAsnArgValGlnGlnProValThrAlaThrAspSerPheArgThr 20
|||||
DB 46117 ATGCACATCAGCAGCGGTCACAAACCGCTGACGCGAGATAGCTTTCGACA 46058
QY 21 AlaSerAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnIleArg 40
|||||
DB 46057 GCCTCGACGCGCTCTTCTCAGCTCTGTCGATCTGTCCAGTCCGATCGCAACGC 45998
```

```
QY 41 GluIleAsnAlaIleAlaAspTyrIleuThrAspHisValPheAlaIleHisIleuPro 60
|||||
DB 45997 GAGATTAATGCGATTGCTCCGATTACCTGACAGATCATGTGTTGCTCGCATTAACCTGCCG 45938
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
CGGGCGGATTCGGCTCATGCGCAAGCTGACAGTTGACGATACAAATCGCAGATCATGCGG 45878
QY 81 LeuIleGluThrArgAlaSerArgIleuHisPheGluGlyIleuThrProAlaThrIleAla 100
CTGATCGAAGCGGGCCCGCCCGCTCGCATTCGAAAGGGGAAACCCCGCAACCATCGGCC 45818
QY 101 AspThrPheAlaIleValIleGluIleuAspArgIleuAlaIleThrThrSerGlyAlaIleu 120
GACACCTTGCGCAAGCGGAAAGGCTGACCGCATTTGGCGACGATCATCGAGCGGCTTG 45758
QY 121 ArgAlaThrProPheAlaMetAlaSerIleuGlnTyrMetGlnProAlaIleAsnLys 140
CGGGCGACGCCCTTTGCCATGGCTGCTGTTGCTTCACTACATGACGCTCGATCAACAAG 45698
QY 141 GlyAspTyrLeuProAlaProIleuLysProIleuThrProIleuLysSerGlyAlaIleuSer 160
GGCGATTGGCTGCGCGCTCGCTCCGCAAAACCGCTGACCCCGCTCATTTCCGAGCGCTGCG 45638
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyLysPheuHisTyr 180
GGCGCATGAGCAGGATGGGACCAAGATGAGACCGCGCGCATATTCGATTCGCAATTC 45578
QY 181 LeuSerAlaSerProAspArgIleuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
CTGAGCGCTCGCGGACAGGCTCCACGATGGCATGGCCGCTTGGTGAAGCGCCACTCG 45518
QY 201 ProSerLeuAlaArgGlnValIleuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
CCAAAGCTTGTCTCGACAGGTTCTGACAGCGGGGGTTCGGTTCAGACACTCGCGCGCC 45458
QY 221 AsnAlaValArgThrValIleuAlaProAlaIleuAlaSerArgProAlaValGlnGlyAla 240
AACGCCGTACGATCGCTATTTGGCTCCGGCACTGGCTCCAGACCCCGCGCGGGTGGCT 45398
QY 241 ValAspLeuGlyValSerMetAlaGlyIleuAlaAlaAsnAlaGlyPheGlyAsnArg 260
GTGACCTTGTTGTATCATGATGGCGGGTGTGTGGTCCGCAACGACGCTTGGCAACCGC 45338
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValIleuGlyLeuLys 280
CTGCTCAGTGTGCGAGTGGCGGTGATCACACGCGGCGGTGCATTAAGTGTGCTTGAAG 45278
QY 281 AspLysGluProLysAlaGlnIleuSerGluGlnAsnAspTyrIleuGlnAlaTyrLysAla 300
GATTAAGAGCCCAAGGCTCAACTGACGAGCAAGAAACAGCTGGCTCAAGGCTTATTAAGCA 45218
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
ATCAAAATCGGCCACGATCTACGCGGTGCGGCTCAACAGCTGGCAAGGAGATGGCGGTTCG 45158
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
CCACTGATATGCGACCGACCAAGTGGTGGCGTAAGAAAGCTGTGTCAGGCTCGACAC 45098
QY 341 LeuThrGlnAsnGlyLeuAlaIleuAlaGlyIleuPheAlaGlyValGlyLysLeuGlnGlu 360
CTGACCCAAACGCTTGGCTGGCGGTGGCTTGGCAGGGGTAGGCAAGTTGCGACGAG 45038
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnIleuThrAsn 380
ATGGCAGCAAGAAATATCAACCGACCGGGGACCAAGCGCGGTTCAGTCACTGACCAAC 44978
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrPheThrAlaAlaIleuThrThrAspPro 400
CTGGCAGGTTCGCGACCGCTTTCGACAGCTGACACCGCGCGCTGACCAACGATCC 44918
DB 44977
```

QY 401 AlAValLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 44917 GCGGGAAGAAAAAGCCGAGCGCTTCATACAGGACAGCGTGAATCATGCTCATCCAGTACC 44858
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 44857 ACAGGCTACGTAGCGGACGACGCGTCAACCTGGCGAAGACGTCGAACACATGGCGGG 44798
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 44797 GAGGCGATCACCCATACCGGCGCCAGCTTGCCTAATACGCTCATACCTGGCTACAGCC 44738
QY 461 ProAlaIleArgGluAlaAspIleGluGluGlyThrAlaAlaSerProSerGluIlePro 480
Db 44737 CCGGCTCGGAGCTGATATAGAGAGGCGGCGACGCGCTTCTCCAGTGAATATACCG 44678
QY 481 PheArgProMetArgSer 486
Db 44677 TTTCGGCTATGCGGCTCG 44660
RESULT 2
AF458051 1834 bp DNA linear BCT 02-MAR-2002
LOCUS AF458051
DEFINITION Pseudomonas syringae pv. maculicola type III effector HopPtoAlpma
gene, complete cds.
ACCESSION AF458051
VERSION AF458051.1 GI:19071505
KEYWORDS
SOURCE Pseudomonas syringae pv. maculicola.
ORGANISM Bacteria; Proteobacteria; gamma subdivision: Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 1834)
AUTHORS Gutman,D.S., Vinatzer,B.A., Sarkar,S.F., Ranall,M.V., Kettler,G.
and Greenberg,J.T.
TITLE A functional screen for the type III (Hrp) secretome of the plant
pathogen Pseudomonas syringae
JOURNAL Science 295 (5560), 1722-1726 (2002)
MEDLINE 21862332
PubMed 11872842
REFERENCE 2 (bases 1 to 1834)
AUTHORS Vinatzer,B.A., Gutman,D.S., Sarkar,S.F., Ranall,M.V. and
Greenberg,J.T.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2001) MCB, The university of Chicago, 1103 E
57th Street, Chicago, IL 60637, USA
FEATURES
source location/Qualifiers
1. 1834
/organism="Pseudomonas syringae pv. maculicola"
/strain="ES4326"
/db_xref="taxon:59511"
/note="pathovar: maculicola"
91. 116
/note="hrp box"
280. 1734
/note="similar to Pseudomonas syringae pv. tomato HopPtoAl
encoded by GenBank Accession Number AF23204"
/codon_start=1
/transl_table=1
/product="type III effector HopPtoAlpma"
/protein_id="AAL84253.1"
/db_xref="GI:19071506"
/translation="MYINRSISOSISSTESFHSQSVASSRVLSAEQTKLNDIT
HYLDHYFAHRLPVNDINVDQVYVHAHNVEIKGISARLRILDEGFTALSIGETFA
KAEKFDMAASAASALRATPRPAASVLYOYMPALINKGMLTPPLKPLAPVYSGALSY
MDQVGTGVMNPAIDDLHTLSTAPERLHDMAASVRRHPGVWRPADIDGLAVQYTTAR
NALRTVLAAPALASRPAAVAGADIVSAAGSLVANAQSERMHTVQARHDLGAGFVLG
IKDQPRADLSEETDWLDAYRAIKSASVSGALNAGKVAAGLPLDVAIDGKAVRSIV
SATSLONGVLVAGAGVGVKLOEMATKNIHPATKAASQUTFNLAGAPVESMTTVA
AVATDPVAKAESFLIDPVKTGVNGSTALADKTKLAKAIDSAERIATPTGASLRD
TLRRRTVREPIDEGGVAAGTSGPYAHPFEAMRP"

Alignment Scores:
Pred. No.: 1.52e-82 Length: 1834
Score: 1653.00 Matches: 337
Percent Similarity: 79.30% Conserved: 50
Best Local Similarity: 69.06% Mismatches: 93
Query Match: 68.53% Indels: 8
DB: 1 Gaps: 3
US-09-825-414-7 (1-486) x AF458051 (1-1834)
QY 1 MethisIleAsnArgAlaGlnInProProValThrAlaThrAspSerPheArgThr 20
Db 280 ATGTACATCAACCGCTCCATCAGTTCAGTCCAGTATGGCAGCGAAGCTTCATCT 339
QY 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerAspGlnArg 40
Db 340 GCGCAG-----TCTGTGCCCTCCAGTTCGTCAGAGGCTCAGTACGAGCCGACAGC 393
QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
Db 394 AACTGAAATGACATCACCCTGATCCTGACAGATCATGTATTCGCCGACACAGTTACCG 453
QY 61 ProAlaAspSerAlaAspGlyGlnAlaValaLysValHisAsnAlaGlnIleThrAla 80
Db 453 GTAAAGATCAACACGTTGATCGATGCTTCATGCCACAAACGTCAGATCGGCAAA 513
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
Db 514 ATAAATCAGCGCCCGTCCCTACGCTTGTGATGAGGCGGAACCGCTCTCAGCATTTGGC 573
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
Db 574 GAACCTTTGCCAAGCAGAGAAATTGACAGCATGCAAGGCGGCTCGGCGCTTTA 633
QY 121 ArgAlaThrProPheAlaMetLaseLeuLeuGlnIleTyrMetGlnProAlaIleAsnLys 140
Db 634 AGGCAACACCGCTTCCGCGGCGTCCGTCGTCATCATCATCAGCCCTGCAATCAACAG 693
QY 141 GlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyValaLeuSer 160
Db 694 GCGCAGCTGTTACCGACCGCGCTCAAGCCCTTGGCCCGCTGCTGTCGCGGCCCTATCG 753
QY 161 GlyAlaMetAspGlnValGlyThrLysMetLaseAspArgAlaThrGlyAspLeuHisTyr 180
Db 754 GCGCTCATGACAGCGTGGCGACCGCGCTCATGAACCGCGGACCGGAGCTGCATTAC 813
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaIleSerValLysArgHisSer 200
Db 814 CTGACACTCGCGCGGAAACCTGACAGATGCGTGGCTGCTGCTGTAACCGCATAG 873
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValaIleAlaGlnThrTyrSerAlaArg 220
Db 874 CCGGCTCATCGCTCAGCGCTCGGATCTGGGATCGCGCTTCAGACTACAGCGCGGA 933
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnIleAla 240
Db 934 AATGCGCTGCGTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 993
QY 241 ValAspLeuGlyValaSerMetAlaGlyLysLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 994 GTAGACATCAGCGTCTCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1053
QY 261 LeuLeuSerValGlnSerAspArgHisGlnArgGlyValaLeuValLeuGlyLysLys 280
Db 1054 ATGCTACGCTGCGACGCGCTGATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1113
QY 281 AspLysGluProLysAlaGlnLeuSerGluGlnAsnAspTyrPleuGlnAlaTyrLysAla 300
Db 1114 GACAAACCAACCAAGCAGACTTGAAGTGAAGAAACCAACCTGCTGCTGCTGCTGCTGCTG 1173
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320


```
/product="PUTATIVE TRANSMEMBRANE PROTEIN"
/protein_id="CAD18268.1"
/db_xref="GI:17431591"
/translation="MNDAISHETITISCFLEHORPTVEIAERENRDEAOYRQALRAL
ADNHCYVLPSPSDEWIAHPSAAPTCTCVKSGNKGNCAMCALGIVHLAGSATI
ETRGALIDDHVIIRIENGRLIDTDYVVFPIPMKHAMDNVITCSVQLPFDEKQVDE
WCASRGIPKGDVRLPEQVWNFVEMVYAHADADWTMSVREAIELFAHHLSGPWITI
AAEAAKRF"
complement(3188..3595)
/gene="Rsp1118"
/note="RS05501"
complement(3188..3595)
/gene="Rsp1118"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18269.1"
/db_xref="GI:17431592"
/translation="MFDTKIAIVRDLDATWOKLNTVAFMNSGITGASPEIMGEAYRD
AADNTYHSLCQPVIVLSGDETLKAVRRLERKVPGLYIEDWFSTGHDAANRAVE
AEHGDSARVYGLALREKKITVDKTKARRHP"
complement(3673..4161)
/gene="Rsp1119"
/note="RS05500"
complement(3673..4161)
/gene="Rsp1119"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18270.1"
/db_xref="GI:17431593"
/translation="MGSEVGEVPCPIARLTVAVGDRVTLLIREPLGYKRFEEFOA
OTGMSHLSLSTRKRLKLEADGIVVRHLISDRPRRHYRLTSGLDPLPILLSKSGEK
WGFNAKTAAPALITTHRCGHETGLKVCAPCDEPFGKDAITVILGASFKAERQARD
GA"
4228..4725
/gene="Rsp1120"
/note="RS05499"
4228..4725
/gene="Rsp1120"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18271.1"
/db_xref="GI:17431594"
/translation="WPATIRHPRGAPRYLDFKXGDLTETSGEAVTMDILSFAEOY
DPQPHLDEIARDVITGELVSGWOTLAVTMRLLDARLLGTPIVGAEELDRIFHA
PMRPDGVARAEVIAIRPSKSRPDRCFMDARVTTTNSGVTLVTFQMSLVVPRPDE
OKVAF"
complement(5017..5610)
/gene="Rsp1121"
/note="RS05498"
gene
```

```
CDs
complement(5017..5610)
/gene="Rsp1121"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD18272.1"
/db_xref="GI:17431595"
/translation="MAEKITRRKPRADAERNRORLLDPAKTAFAEKVSASLEDIARE
AGVIGITLYRRPFRDLIDETVRSDEGSLAEAROSEOPPLEAVROWITLIPVYL
ANKOIRADVLNCFMDCSERCTLSGEVILLETIADLITERAKOSGATIGLAVEPLDLSAV
AGVASFGAELDWEAGARLVEWVAGLRVAGATGKR"
5822..6910
/gene="Rsp1122"
/note="RS05497"
gene

Alignment Scores:
Pred. No.: 6.4e-17 Length: 208050
Score: 508.00 Matches: 171
Percent Similarity: 46.73% Conservative: 72
Best Local Similarity: 32.88% Mismatches: 217
Query Match: 21.06% Indels: 60
DB: 1 Gaps: 14

US-09-825-414-7 (1-486) x AL646083 (1-208050)
QY 3 ILEASARGARGVALGGLNGINPROProValThrAlaThrAspSerPheArgThrAlaSer 22
||| |||||:||||: ||| :||| :|||:
Db 203441 ATTCCGCCAGCGTCTCCGGCATCCCGAGGCGCTGTAAGCGGCCAATCCAGCGCAGCA 203500
QY 23 ASPALASerLeuAlaSerSerValArgSerValSerSerAspGlnArg-Glu11 42
||| ||||| ||| ||| ||| |||
Db 203501 ACAGCGAGTGTAGCCAGTGTCTCCGGCGGCGCTCCATCCGCGCACCAATGCGGAT 203560
QY 42 EASNAIleAlaAspTYrLeuThrAspHisValPheAlaAlaHisIleuProProAl 62
|:: ||| :|||: ||| :|||: |||: |||
Db 203561 CCATGGAATC---GAATTCGCCGCGCGGGGGCGTACCGCCGATCCCGCCAGCCGGGCG 203617
QY 62 AASP-----SerIAspGlyIleAlaVal-AspV 73
||| |||||: |||: |||: |||:
Db 203618 AGACCGCGCGTGCATCCAGCATCCCGCACCTGATGCCATAGCCAGCAGCTGCCGATA 203677
QY 73 al-----HisAsnAlaGlnIleThrAlaLeuIleGluThrA 85
:: |||||: |||||: |||||: |||
Db 203678 TTTTCAGCGCGCATTTTGACGGGTACCAACGATGCCCAATTGAGATCTGATCCAGCAGC 203737
QY 85 RGAISerArgIleuHisPheGlyGluThrProAlaThrIleAlaAspThrPheAla 105
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203738 GCGCAGAACGGCTCCAAAGCGATGGGGGAAACCGCGCAACGCTCCCGCCGTAAGTGGCA 203797
QY 145 YSALGluIleuSerPheArgLeuAlaThrThrThSerGlyValaLeuArgAlaThrProp 125
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203798 AGGGCGCCACACCGGAGCCGGCTGGCCAGACCGACGTAAGCTGTGCGGTGGCGCGGT 203857
QY 125 heAlaMetAlaSerLeuLeuGlnTYrMetGlnProAlaIleAsnIleuSerGlyAlaPleu 145
||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
Db 203848 TCGGCGATTGCTCGCGCTGTTCGACGTGAAGCAGCAGCACTGACGGCTTC----- 203907
QY 145 roAlaProLeuIleuSerProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 165
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 203908 --GCCAAGACGACGGCCCGAGTCGGCGGACCGCTGGGTGGCGGACGGCGGATG 203965
QY 165 InValGlyThrIleuMetAlaSerArgAlaThrGlyAspLeuHisIleuSerAlaSerP 185
||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: ||| :|||: |||
Db 203966 CCTTGGGGGCGACCTGCTGGGCAAGGCCACGACACACAAAGGCTGGGCGGCTGGC 204025
QY 185 roAspArgLeuHisAspAlaMetAlaIleSerValIleArgHisSerProSerLeuAla 205
```

[illegible]

Qy	462	laargtguablaapligluglugly-----glythralaalaaserp	476
		:::: ::::	
Db	5904	CCGCGCCGCGCTGTTCGGGGGCTGTGTCAGCGCTGCTCCACCGTCCGCGCCG	5845
Qy	476	rosereululeprophetaipro	483
		:::	
Db	5844	TTGCCCGCATGCGCCGCCGCC	5822
RESULT 5			
MTV023/c			
LOCUS	MTV023	47852 bp	DNA linear BCT 03-AUG-2001
DEFINITION	Mycobacterium tuberculosis H37Rv complete genome; segment 148/162.		
ACCESSION	AL022022 AL123456		
VERSION	AL022022.1 GI:3261554		
KEYWORDS			
SOURCE	Mycobacterium tuberculosis H37Rv.		
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.		
REFERENCE	1 (bases 1 to 47852)		
AUTHORS	Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E., Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squares,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and Barrrell,B.G.		
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
JOURNAL	Nature 393 (6685), 537-544 (1998)		
MEDLINE	98295987		
PUBMED	9634230		
REFERENCE	2 (bases 1 to 47852)		
AUTHORS	Parkhill,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk		
COMMENT	On Jun 27, 1998 this sequence version replaced gi:2924430.		
	Notes:		
	Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.		
	(URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.		
	Gene prediction was based on a Hidden Markov Model of TB genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.		
	CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.		
FEATURES	Location/Qualifiers		
source	1. 47852		
	/organism="Mycobacterium tuberculosis H37Rv"		
	/strain="H37Rv"		
	/db_xref="taxon:83332"		
	<1. .18		
source	/organism="Mycobacterium tuberculosis H37Rv"		
	/strain="H37Rv"		
	/db_xref="taxon:83332"		
	/clone="Y13B12"		
	47752. .>47852		
source	/organism="Mycobacterium tuberculosis H37Rv"		
	/strain="H37Rv"		
	misc_feature		
	<1. .47852		
	/note="fragment designated v023. Does not represent a physical clone"		
	complement(1. .1695)		
	/gene="Rv3494c"		
	complement(1. .1695)		
	/gene="Rv3494c"		
	/note="Rv3494c. (MTV023.01c), len: 564. Unknown Pro-rich protein similar to several Mycobacterium tuberculosis proteins e.g. MTC128.14 (515 aa), MTCY19H5.28c (516 aa)and (MTV051.09). Has hydrophobic stretch, possibly signal peptide at N-terminus. FASTA scores: 297050 MTC128.14 (515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25; 297182 MTY19H5.28 (516 aa) identity in 559 aa overlap: 297182 MTY19H5.28 (516 aa) opt: 979 z-score:567.7 E(): 4.1e-24; 33.58 identity in 555 aa overlap. TParse score is 0.897"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein Rv3494c"		
	/protein_id="CA17731.1"		
	/db_xref="GI:2924431"		
	/db_xref="SPTREMBL:O53539"		
	/translation="MIDRLAKIQLSIFAVITVITLSVAIFLRLPATFGITGVSA DVAGGKLYKANAVYRGVAVGRVSGVLPNGVTAHRLNSGTAPSNVATVRSV AIGREYDILVPENSSSKLRNGRFIQGNRNIQGVADLRQAETLGSIGDPTRE LIEHFINTNGAPPLARLIESARLIVDEANNTYQVSQIDQACPFIQAOIRACGDI KSLAAGLARFTWOLRAADPRLDLPADDAIDENATVFSGRPSFPALASLANLGR VGVYHKSIEQLVVEPALFAIITTSAGGVQDEGAKLDFKIDLPDCMGWGLPVP LVRSPADSVREIPRDYMKCTQANDPSEYRGARNYPCQEPKRAPYQLCDPBGV PVGTNPMRGPIPIYGTGYTDGNILPKNKFPYIPGADPDPVPTVGGPPQVGGV PAPHQAPAPPPNDNGPPTPTSMPCYPPPEQVYYPATIPPPPPPCGPPCP APGPQASGPATYITTDLSGAFADPAGCTGTFAPGMTGASSAENWVDLMRDPRL"		
	complement(1703. .1707)		
	/note="possible RBS for Rv3494c"		
	complement(1706. .2860)		
	/gene="1prn"		
	complement(1706. .2860)		
	/gene="1prn"		
	/note="RV3495c. (MTV023.02c), len: 384. 1prn, similar to Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and MTCY19H5.29 (402 aa) and (MTV051.08). Probably 11proteins, contains possible signal sequence and appropriately positioned PS00013 Prokaryotic membrane 11proteins lipid attachment site. FASTA scores: 297050 MTC128.13 (390 aa) opt:653 z-score: 762.1 E(): 0; 33.6% identity in 363 aa overlap; and 297182 MTY19H5.29 (402 aa) opt: 572 z-score: 667.9 E(): 1.1e-29; 31.8% identity in 362 aa overlap. TParse score is 0.897"		
	/codon_start=1		
	/transl_table=11		
	/product="1prn"		
	/protein_id="CA17732.1"		
	/db_xref="GI:2924432"		
	/db_xref="SPTREMBL:O53540"		
	/translation="NKRITWLRILITFASALLIAGCGFGLNSLPIPTAGHCEGAYSV IVEMADVATLIPONSVMVDVYVGSVAIVAVQRDGSEFYAVKLDLKNVLLPNNAV AKVSGTSLGSLHVELAPPTDPTGRLVVDGSEIRLEANTDPTPEEVSALGVVYK GNVGALSEIIDEITHOAVAGROQFVNLPVRLAELTAGNROVHDIIDLDGLNRVSAI LARQKDNIGRALDITLPDAVRYVNONRDIIVAFALKRLTWTSHVLAETVDFEEDL KDLISIVAILANDRKDPYTSQDLITTFPPPNNGICQANRGDILNFTTFPDLTKRIGE TFFTYATFDPNNAHDELTINPDLFIELANLSGAADPFKIPETASGO"		
	complement(2798. .2830)		
	/gene="1prn"		
	/note="PS00013 Prokaryotic membrane 11proteins lipid attachment site"		
	complement(2857. .4212)		
	/gene="Rv3496c"		
	complement(2857. .4212)		
	/gene="Rv3496c"		
	/note="RV3496c. (MTV023.03c), len: 451. Unknown but		

similar to Mycobacterium tuberculosis proteins MTC128.12 (530 aa), MTC119H5.30c (508 aa) and MYV051.07. Hydrophobic region at N-terminus. FASTA scores: 2970501MTC128.12 (530 aa) opt: 838 z-score: 711.0 E(): 4.3e-33; 35.1% identity in 473 aa overlap; and 2971821MTC119H5.30 (508 aa) opt: 821 z-score: 697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap. Tbpase score is 0.891"

```

/codon_start=1
/transl_table=11
/product="hypothetical protein RV3496c"
/protein_id="CA117733.1"
/db_xref="GI:2924433"
/transl_table="EMBL:053541"
/translation="MMGRVAMLSGLRATYATVLAVALVAGVYVSSSTGKRTITVG
YFSAAGLPEVDYRATVGVPGVGTIDIRPSSSDYKTKMSKDYVPVQVAYMSRRP
LVARFQILVPYITGGAVLPDKNRQILDLTRAPVDEKESLRLTLADSLVAGSELQ
GLPLAALIQADALIDLDGSDVLSHNLDELQAVNRGDSKGLFFGVKRLQVLVAGELQ
SDEQIYVPGFASVYLAVDSSANIDQPTGLNLALSDIRFLKNNSTLITVNDQI
NDKQTLSDQSENTEVQVLAAGGTLITLIPNPAAGTGLTSLSPNPAVQPCGSG
SEFDIAGSGADYRRAEIRKELGVRLRLVNPDPHFHLNLTITVKQIITYDPR
ATPEKSEVPELDTWPAAGGAVGNRQDLSLVPAPRQAPAPRAGAPGSHGGGG
G"
/complement(2867..2871)
/gene="RV3496c"
/note="possible RBS for RV3495c"
/complement(4209..5282)
/gene="RV3497c"
/complement(4209..5282)
/gene="RV3497c"
/note="RV3497c, (MYV023.04c), len: 357. Unknown but
similar to Mycobacterium tuberculosis proteins MTC119H5.31
(481 aa), MTC128.11, (515 aa) and MYV051.06. Hydrophobic
region at N-terminus. FASTA scores: 2971821MTC119H5.31
(481 aa) opt: 611 z-score: 667.7 E(): 1.1e-29; 37.5%
identity in 332a overlap; and 2970501MTC128.11 (515 aa)
opt: 587 z-score: 641.2 E(): 3.3e-28; 30.1% identity in
335 aa overlap. Tbpase score is 0.889"
/codon_start=1
/transl_table=11
/product="hypothetical protein RV3497c"
/protein_id="CA117734.1"
/db_xref="GI:2924434"
/db_xref="SPRMBL:053542"
/transl_table="MLNKRKSHHEHDDRLRTGILFGLVILCVLITAFGYSGLPFWPOGQ
KYVAPFLIDAGITVPGVYVSSLSKLVGASVLAAGNSAKYFSDRSIIVGDGSLAAI
LRTVTLIGERSIATVPSVSGSKSTTIFLSTPTTPTATGLDGLNADLRPEQDRA
LNVETQALHATPAVGRVAVGLTSLRSLNRREDAIOGLLAARKEVTSLEERAEQAL
LIVTEGQNLPEALDARAALATSLISGIDVAAOISFVADNKKRGPALSTNLIVLAN
INERDQITFEKRLPLVATLTGEEVGSGRGVNYSVLPRLPATVFDLVPQKGL
PDSLADILRGTQDRMLIRKSP"
/complement(5272..6324)
/gene="RV3498c"
/complement(5272..6324)
/gene="RV3498c"
/translation="MYV023.05c", len: 350. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.10
(346 aa), MTC119H5.32c (346 aa) and MYV051.05. Hydrophobic
region at N-terminus. FASTA scores: 2970501MTC128.10 (346
aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in
340 aa overlap; and 2971821MTC119H5.32 Mycobacterium
tuberculosiscom1 (275 aa) opt: 769 z-score: 820.1 E(): 0;
47.0% identity in 249 aa overlap. Tbpase score is 0.878"
/codon_start=1
/transl_table=11
```

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:

Length:	47852
Matches:	95
Conservative:	43
Mismatches:	133
Indels:	98
Gaps:	14

US-09-825-414-7 (1-486) x MTW023 (1-47852)	
QY	145 ProAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMeta 164
Db	31455 CCGGGCCG-----CCGCTGGCCGGCGTGTAGAGAGAGAGCTGCCGGCGCCG 31402Z
QY	164 spGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlas 184
Db	31401 GTGCCCGCGCACCGCGCTTGGCCGCCGCTTGCCAGCGGT----- 3136Z
QY	184 erProAspArgLeuHisAspAlaMetAlaLysSerValLysArgHisSerProSerLeuA 204
Db	31363 -----GCCGTAGCGCGCCACCGCGCCGCCGCCGCCGCCGCCG----- 31328Z
QY	204 IaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValA 224
Db	31327 -----CGCCCCACCGGACCCACC----- 31310Z
QY	224 rgtThrValLeuAlaProAlaLeuAlaLysSerArgProAlaValGlnGlyAlaValAspLeuG 244
Db	31309 -----GCTGGCCGCTTGAACCTCTTGCACCGCGCTTGCACCGCGGAGTGGGTG 3126Z
QY	244 lValSerMetAlaGlyGlyLeuAlaAlaAlaAlaGlyHeGlyAsnArg----- 260
Db	31260 GGCCTGCT--GCTCCGGCGCGCGCCGACCGCGCAGACTTCCCGTCCGCGCGTGGCCG 31204Z
QY	261 --LeuLeuSerValGlnSerArgAspHisGlnArgGlyValAlaValLeuGlyLeuL 280
Db	31203 CCTTGGCCCGCGCGCGACCGACCGATACCGTGGCGCGCGCGCACCACT----- 31154Z
QY	280 yAspLysGluProLysAlaGlnLeuSerGlnGluAsnAspThrLeuGlnAlaTyrLysA 300
Db	31154 ----- 31154Z
QY	300 IaIleYserAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL 320
Db	31153 -----GCCGCTCTGCCCGCGCGGCGCGCTTGCACCGCGCAGCCGCGTCCGCGCC 31102Z
QY	320 euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAlas 339
Db	31101 CCACGGTCTCTCCAGCCGTGGCGCGCCGCGAGCGCCCGCTTGGCGCGCTGTCGCGCGC 31042Z
QY	339 erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly--GlyPheAlaGlyValGlyLys 358
Db	31041 GTACGCTTGGCCCGACCGGTACCGCGCTTGCCTGCCCGCGCACACCGCACCGGCT 3098Z
QY	358 euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGln 378
Db	30985 -----GTTTCCGCGCTTCCCGCACCCGCTTCCGCGCTATTTCCGCTA 30937Z
QY	378 euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyThrThrAlaAlaLeuThrT 398
Db	30936 CCACCGGCGCTTCCGCGCACCGCGCGGATGGCGCGTTCGCCGCTTGCACCGGATCT 30877Z
QY	398 hrAspProValLysLysLysAlaGlySerPheIleGlnAspThrValLysSerThr--- 416
Db	30876 CCGGTCCGCGCGCGCGCGCGCGCGATGCCGATACCGCACCGACCGCGGTTCCACCGTTG 30812Z
QY	417 -----AlaSerSerThrThrGlyTyrValA 425
Db	30816 CCGCGCGCGCGCGCGGTTCGCCATGCCCGCGCGCGCGCGCGCGCGCGCGGAGAT 30757Z
QY	425 IaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGlnAlaIle--- 443
Db	30756 CCGAGGCGACACGCGCGCGCATTCGCGCGTGGCGCGCATTCGCGCACGCGTACCGTTGGCG 30697Z
QY	444 -----ThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA 462
Db	30696 CCGGAGGTGCACCGCGCGCGC-----CGGCGCGCGGTCTCCGCGCTTCCGCGCGGTGCGCG 30644Z
QY	462 IaArgGlnAlaAspIleGlnGlyGly-----GlyThrAlaAlaSerP 476
Db	30639 CCGCGCGCGCGCGCTTGGCGCGCGGTGCTGTGTAGACCGCTGCTTCACCGTCCGCGCGCT 30580Z

[illegible]

```

/product-conserved hypothetical protein"
/db_protein_id="AAK23001.1"
/translation="MODRIYAQVQEPNPAPKLRGASGDLVARKVYCMQOMRPPTARGEYV
STREWSGYATDVYVYTGELGCLPGLDYLTVGRHAPRTIENAAEGLGFGOGSIVSH
AAQPCIRNMYGQDFNPNSQAAFAAEWRLSGAEKLYIDEAEFCNRDLDPPQETIGH
GIVTITSDNRHVLDPYIRKKLRPGCVLTISTVTLPEWSSAPRTIHLKTHADVMGAP
GGCIGTQDTAAALNFIDQFLALDYLAKVNGVAERLKQVAAQDRKTLAHEYMNDDMPH
MYFADENWMLGDKAFVAGSAGVAHLDNLIDLNMVPOQTEFMRDVSLSRETRIFCNNA
OEFRRBYRWYGVALLGHEQVAAHREERYVTVQADLPKRVGVIGEATLNDVYQPL
LEFILLDHPKSHVLELETILNASKSLHRGQSSALITLLGGLGTTTAAADVSASKAR
TIALNKTKTARRSLSGVDINAFASPTGGGTTSARFQLPFLSLNRQNGKTAEDNASFAY
QVLYAQKAIKDGATLQGDINDYITELRQAQTVGHAYLPIMNLGV"
/gene="CC1018"
/complement(3751..4500)
/gene="CC1018"
/complement(3751..4500)
/gene="CC1018"
/complement(3751..4500)
/note="Identified by Glimmer2; putative"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/db_xref="GI:13422310"
/translation="MPLLDILAVTVFACWLLXEPMLKRIGESGAVLNTDMYIIRRMW
NMEMAARETALDDGLQGLGHAINSFSSNLILIAAAGVLGCGDALKSDEBLAVLV
AKTTPMPQIKCLVLYAGVGLDFTWISIRQNNYCLAIGATPMAPNVLVEASAPAA
AGGILNPALASFAFAGRAYVFPALAAVWVPLPLTATIGAMILLIMRORSSASAVA
VVKRYOLRQPVHGRPHLSRTSTPEDRKI"
/complement(4542..5975)
/gene="CC1019"
/complement(4542..5975)
/gene="CC1019"
/complement(4542..5975)
/note="Identified by match to protein family HMM"
/codon_start=1
/translation="major facilitator family transporter"
/product="major facilitator family transporter"
/db_protein_id="AAK23003.1"
/db_xref="GI:13422311"
/translation="MATHTAEMLTPVAPARAILTQESAGNLVEMDMPAFVAAFLTPAA
PAFFPRGDDTVQLQAAVFFLGFVAPRPGIAMTIGLADYADGRRAALSVSLMACAA
FYIATIPDPTQIGLMPALILFPRVLDQGSVGGEGASAVYNSEMGKARGFMSSRH
YLLTILAGQLADLILDLRTLEDEALKEWAMGMPVPIGALLAVVWYIRRGLEESYV
FKRPAESEPILSRQIISAGAFSLTASAGVGMGPKLAKAGQYLGAFVLFPSLIL
VPLRPPRESLITIMLTAGGSLTFPVYVYTYMOKPLVNTAGFSKAQSEISALSILAF
MIMQPLAGLSDEPKGPKMLITAFGGVGLTIPMIMATISQNSGVALLILNGVAPQ
SCYTATISAVYKAMFPAEIRALCVALLPYALANLVFGSTAEVALLAKRHENDESTFYV
VAAVMTIGLSTIILDTGRHSLIHD"
/gene="CC1020"
/complement(6053..6442)
/gene="CC1020"
/complement(6053..6442)
/gene="CC1020"
/complement(6053..6442)
/note="Identified by Glimmer2; putative"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/db_protein_id="AAK23004.1"
/db_xref="GI:13422312"
/translation="MKASISKQNDVSLVREALEGERTRRTPGGKAGLEPRHQNODDQD
HBDQAAAGAGETLAPRAVPPGGRAQDQDQKNNENAEHGRPVPEAPASATVAREVM
GSEFAMVLVGSPEKGLSKRRLOPCGS"
/complement(6408..6800)
/gene="CC1021"
/complement(6408..6800)
/gene="CC1021"
/complement(6408..6800)
/note="Identified by Glimmer2; putative"
/codon_start=1
/translation="conserved hypothetical protein"
/product="conserved hypothetical protein"
/db_protein_id="AAK23005.1"
/db_xref="GI:13422313"
/translation="MOMSEBPQSPSVLTGLKRGRLRHRCNCSDGRLYRRYLKVPYC
EACGHEGLGAYPDADGQAVYLTIVGHIIILAPMLFPPMIWASPMLVVPLVLTATLT
LTLTPRVKCAVITDVAIKRKEEDPS"

```



```

Db 12558 GTCGTCAGCCCTCGCGCTTCCATGTGCAACCGCTGGATGCCCTGGATCAGCGCTGGT 12499
Oy 302 -----LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLy 315
Db 12498 CATCGCATCGGCTTGGGGGGGGCGCCCTCGATCAGTGGACGGCCCTTGGGCTTGAA 12439
Oy 315 sArGMeLaGlyLeuProLeuAspMeLaIaThrAspAlaMeGlyAlaValaArgSerLe 335
Db 12438 CAGGGCGGGGGG-----CGCACGGCGGGCCAGCAGGGCGGGCTT 12397
Oy 335 vAlaSerAlaSerLeuThrGlnAsnGlyLeuAlaLeuAla-----GlyGI 351
Db 12396 GCGGGCCACTCTCTCGACCGCGCGACAGGGGCTGTCTATGAGCCCGACGGCGGGG 12337
Oy 351 yPheaIa---GlyValGlyLysLeuGlnLeuMeLaIaThrLysAsnIleThrAsp----- 368
Db 12336 CTGGTCTTGAGAGATCGGCACCGCCGATCAGCCAGACACCGCCGAGATCTCGCCCTGCTT 12277
Oy 369 -----ProAlaThrLysAlaAlaVala-----Se 376
Db 12276 CCAGCTGATCGGGCCGTCGCGCTTCCGCGTGACAGCGACGGTGGTGGTGGCGGCG 12217
Oy 376 rGlnLeuThrAsnLeuAlaGlySerAlaAlaValaPheaIaGlyTyrThrThrAlaAlaLe 396
Db 12216 GCGGTCGACGGGCTTG-----CCGGCCAGACCTTTTCCAGCTGGCGACATATATGCC 12163
Oy 396 uThrThrAspProAlaValaLysLysAlaGlySerPheIleGlnAspThrVala----- 413
Db 12162 CACACCACTGGAGAGTCGGCCCGCCCTCCATTCGCTCGCGCCGATTCGCGCTCATGGG 12103
Oy 414 -----LysSerThrAlaSerSerThrThrGlyTyrValaAlaAspGlnThrVally 430
Db 12102 ACAGTATCGCTTACCGCGCGCGCGCCGCGCCAGCGCT-----GAAGCTCAGCG 12055
Oy 430 sIeuAlaLysThrValLysAspMeGlyGlyAlaAlaIleThrHisThrGlyAlaSerLe 450
Db 12054 GCTTGGCCGGGTCGCGCATGCTGTGGGGGGGGGCGTCTCTCAGCGGGGCTTCGTCCTT 11995
Oy 450 uArgAsnThrValaAsnAsnLeuArgLnaGly-----ProAlaArgLnaAlaAspIle 467
Db 11994 GAGGGCCCGCGCGCTATGCGCGGGCGCAGGGTGTGCACAGGGGCTCGGGC-GCGACCT 11936
Oy 467 eGluGlu-----GlyGlyThrAlaAlaAlaSerProSerGluIleProPheArg 482
Db 11935 TGAAGGTCAGATGCGGCTGTGGCGGCGATGTGCCGAGGTGGAAGACCCCC---CG 11879
Oy 482 gPromeT 484
Db 11878 CCCGCTC 11872

```

```

RESULT 7
SC4A7 42526 bp DNA linear BCT 12-MAY-2002
LOCUS Streptomyces coelicolor cosmid 4A7.
DEFINITION AL134423 AL645882
ACCESSION AL134423.2 GI:20520774
VERSION 1
KEYWORDS

```

```

SOURCE
ORGANISM Streptomyces coelicolor A3(2).
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomycinales; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 42526)
AUTHORS Redenbach,M., Kiseer,H.M., Denapalte,D., Eichner,A., Cullum,J.,

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Kinashi,H. and Hopwood,D.A.
A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
Mol. Microbiol. 21 (1), 77-96 (1996)
97000351
8843436
2 (bases 1 to 42526)
Seeger,K.J. and Harris,D.
Unpublished
3 (bases 1 to 42526)
Cerdano,A.M., Parkhill,J., Barrell,B.G. and Randaiream,M.A.
Direct Submission
Submitted (09-DEC-1999) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK
On May 9, 2002 this sequence version replaced gl:6562773.

COMMENT

Notes:
Streptomyces coelicolor sequencing at the Sanger Centre is funded
by the BBSRC and Beowulf Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S.coelicolor/>) CDS are
numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).
The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous. The length
in codons is given for each CDS.
Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon. Gene
prediction is based on positional base preference in codons using a
specially developed Hidden Markov Model (Krogh et al., Nucleic
Acids Research, 22(42):4768-4778(1994)) and the FramePlot program
of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nhn.go.jp/juncgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg, ttg or (atc)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.
IMPORTANT: This sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions. Cosmid 4A7.
Location/Qualifiers

FEATURES

source

```

1. 42526
/organism="Streptomyces coelicolor A3(2)"
/strain="A3(2)"
/db_xref="taxon:100226"
/clone="cosmid 4A7"
/size=630
637..1320
/gene="SC4A7.02"
/feature="SC02374"
637..1320
/gene="SC4A7.02"
/feature="SC02374"

```

CDS

```

/gene="SC4A7.02", probable transcriptional regulator, len:
227 aa; similar to SW:TCMR_STRGA (EMBL:M80674)
Streptomyces glaucusens tetraacyclic C transcriptional
repressor TcmR, 226 aa; fasta scores: opt: 594 z-score:
727.4 E(): 0; 46.5% identity in 215 aa overlap. Contains
Pfam match to entry PF00440 tetr, Bacterial regulatory
proteins, tetr family and a possible helix-turn-helix
motif at residues 49..70 (+3.84 SD)
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB62706.1"

```


D	10803	GCGGATGCCATCAAGCCGTCCAAGACGGCCGGAAGCGCGGC-----	108444
Oy	148	LeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly	167
D	10845	-----CGTACGGCG	10853
Oy	168	ThrlYsMeMetAspArgAlaThrGlyAlaThrGlyLeuSerAlaSerProAspArg	187
D	10854	GTCGAGTTGGCCGACACC-----GAGCAGCAC-----GCCGCCGCGCTGCCAAC	108955
Oy	188	LeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnVal	207
D	10896	AAGAAGAGAGCGGAGACGGCCGACAAGCCGAGGGGGTGAC--GCGCTCGCGGGTGGCAAC	109555
Oy	208	LeuAspThrGlyAlaAlaValGlnThr-----TyrSerAlaArgAsnAlaVal	223
D	10953	GAGTCACGGGGCTACCCCTTACACACCGCGAGCGCGGGGTGGATGCGCGCACTCGGCG	110121
Oy	224	ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln---GlyAlaValAsp	242
D	11013	GCCGAGTACCCGACCGCC-----AAGCAGCGCATTCAGCTCGGATGCGCGTAT	110633
Oy	243	LeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeu	262
D	11064	GTCACACCGACACTCGGCGCGCGGACTCGCCGTAAGACCGGACAGCTGCTGAAAGAGATC	111223
Oy	263	SerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLys	282
D	11124	GCCGACAG-----CAGCAGGCGCTCGCT-----	11147
Oy	283	GluProLysAlaGlnLeuSerGlnLysN-----	292
D	11148	GAGCGCAGCGCGACGACCGGAGAGAGTGGCGCGCCAGCGCACAGAGCTGGCCAAAGCG	11207
Oy	293	-----AspTrpLeuGlnAlaTyrLys---AlaIleLysSerAlaSerTyrSerGly	308
D	11208	GCGAGCGGTACTCCAAAGCGCGCTTACACACTGGCGCGCAAGCGCTTCGCTACGCGGG	11267
Oy	309	AlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAla	328
D	11268	GAGCGCGGCAACTCTCCAAAGAMACCTTGGGTACTCGCAGAAAGCCTCGTACGCG	11327
Oy	329	MetGlyAlaValArgSerLeuAl-----	336
D	11328	ACCCAGGCGCGACAGTCACTCACCCGCAACATCGCTTACGACACCAAGCGCCACCAAGAG	11387
Oy	337	-----SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGly	351
D	11388	GCGGACAGCGCGACAGCG	11447
Oy	352	PheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThr	371
D	11448	TCCGCGCATGCCGACACCTCTGAGACCCGAGCGCGCGCGCACTGCTGGGACAGCGCGAA	11507
Oy	372	LysAlaAlaValSerLnuLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyr	391
D	11508	CAGCGCGCCGAGAC-----GCCCGGACGCGCGCGCGCACCGCC-----	11546
Oy	392	ThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAsp	411
D	11547	GCCACGAGCGCGCGCGCGCGGAGAGCGCGCCCAAGAGCGCCCAAGATGACCGCGCATGC	11606
Oy	412	ThrValLysSerThrAlaSerSerThrThrGlyValAlaAspGlnThrValLysLeu	431
D	11607	GCTCACACAGCG-----GCCGAGCAGCGCGCGAAGAGAG	11633
Oy	432	AlaLysThrValLysAspMetGlyGlyAlaIleThrHisThrGlyAlaSerLeuArg	451
D	11640	GCAAAGCGCGAACATTCACAAAGGAAGAGGCTGGCGACAGACGCGAGCTCCAT- CGG	11698
Oy	452	AsnThrValAsnAsnLeuArgGlnArgProAlaArgGlnAlaAspIleGluGluGlyGly	471

	OY	472	ThraliaAIA 474 	
	Dn	11750	CCTGAAGCA 11758	
RESULT 8				
AEOI2164	LOCUS			
DEFINITION	Xanthomonas campestris pv. campestris str.	ATCC 33913,	section 72	
ACCESSION	AEOI2164	AE008922		
VERSION	AE012164.1	GI:21111662		
KEYWORDS	Xanthomonas campestris pv. campestris str.	ATCC 33913.		
SOURCE ORGANISM	Bacteria; Proteobacteriota;	gamma subdivision;	Xanthomonas group;	
REFERENCE AUTHORS	(bases 1 to 10275) da Silva,A.C.R., Ferro,J.A., Relnach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergro,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Farla,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madella,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Merck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spicola,L.A.F., Takita,M.A., Tanura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P. Comparison of the genomes of two Xanthomonas pathogens with differing host specificities Nature 417 (6887), 459-463 (2002)			
TITLE	JOURNAL MEDLINE PUBLISHED REFERENCE AUTHORS			
12024217	2 (bases 1 to 10275) da Silva,A.C.R., Ferro,J.A., Relnach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergro,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Farla,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madella,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Merck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spicola,L.A.F., Takita,M.A., Tanura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P. Direct Submission Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil			
FEATURES SOURCE	Location/Qualifiers			
	1..10275 /organism=Xanthomonas campestris pv. campestris str. ATCC 33913"	/strain=ATCC_33913"	/db_xref=ATCC:33913"	/db_xref=taxon:190485"
		/note=pathovar: campestris"	470..559 /gene=XCC0657"	470..559 /gene=XCC0657"

gene

note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA039973.1"
/db_xref="GI:21111663"
/translation="MRLSAPRAGTYIVLDDGASPDGVSLLRQ"
1092..8147
/gene="xada"
/note="xCC0658"
1092..8147
/gene="xada"

CDS

note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="outer membrane protein"
/protein_id="AA039974.1"
/db_xref="GI:21111664"
/translation="MNRIRKVMKSLGVMAVVASLGDSPGVSAALLDRFGDLS
LAALAILGASGIALPLPSASQSEYVERGKASPAHAHTIAGSNAASGAVAGVAGD
SSASQSSAIGRQNALGENAVAIQNSYPIVRQENGKVALGANAGVTGANSVALGAG
STHEDDVYSVSGNGRGCPATRTITNYSAGVNAADANVAALQDAEMVADLITFRK
APGQDLSIGAATDAGAAVADDAANAESGALNAGSNAALADNATSEGAALAVGQN
SAAFYSNVQNGPQSVAAVAGAVDEGDPILITNGVPPDITASGVTAGVAVASASA
DGFASVYAGVYAGAAQSSAFYGVSNATGTAVGTISESGEATGAVGATVAGPDLIP
GLGFVVOIGSEASALGAGALISGTYNTTAVGTISESGEATGAVGATVAGPDLIP
AGPSSMASGELSTALGYSTARGNANVALGANSATPDIVSGVAGAEQRLITVAAQ
GREDDVAINLQNTITAVSDVATASATAPATDGAALAGQSVAGSNAAFNDYSTR
LGSSVSAAGQRTAVGSAKNTTDMATVGRNSTVAVENTILAGNSASGDSGTAVG
GATRTASGATALIGESTANGDSTALDVGSVATRGSTAVAGASVAFGDSAAFGAN
AAAGTATSTALIGANSNAGFERTIVALGASNSGDESLALGVSLSLAGTTAVGNSN
ASIAAATVATFENSSAGADYATRALGDSNASQYFSTAVGSGTANGKATATGYESIGN
GAIISGGLSPFASVAMEGCTAIGTESVAYAGDSTALGANAYAADGTIAVGYANAYOP
RAISGLGGSNAAGDESLALGMEPAQEPQGLALGSGQADAYSTAIQGYATASGATP
AGNSNRVDGATGALIGDSMASGNEFTTGGASVAGSGRATIGAVASGEMDRTVAL
GRESVAADVTALGAMNARAYSDSVATGANNASNSYSLGYANVATGSAASITGG
QSYAAGNSVYALGMOSSNAGSTRSVSLGSGATTPADGVYALGSLTADNTVNSVSGV
SROKTTNVAAGTEGTDVAILDQILNVAIVSTAITARLVAIGTDGTAFFAQGEDATVAGSN
AIADVDYSAPGASQATVAGIAGVAGSASATVAYANASGYNAASGVSATCAFQAO
ASGDVVALIGRSEASGQSTAVGAAAGASGDAFGAGALSVAEGSTALGYASATP
GSSAVAGVAEESVADGTSAAFGFGAEMTSNTALGGYSTASGNSTALGNEFSAQSS
SSVANGGATATGAGYSTAAGGASVAGSYNVAAGVAILGILIPTEASGDFTALGAGMA
AGVNSTLGNPRAEYSGSVALDASVADDPFAYVSGSNNENQITNVAAGCTGTPRA
VNUDLTVAETAOCTSKYFRKASGDDSDACATTEGDBALAGGANSNSNSTAVGA
GAQAAEAAATVAGMDALNSIGGAALGNMAQALNENSVASLADIGATNAGAA
GAISYATATLSEAVASDNOATVAGFRMAASVSAFVSGSBSGRISMSLIGAVAA
SBDYATVAGAGLISGASAVAVGESEFTGCESVAGVGGSTFTPIPARQSTAAAGCG
GGAMATVATYTTAIGNNSYADVAGNALSQASAAVALADNALIGNSRPAIDAGASVVDG
ASATINSTGQGVROYNIGENNAVSQNSYFPRSAENVKALGANAAGTVALGSGS
SRTYANTYNSVSGNGRGCPATRTITNYSDEVATDAYNKQILDALADVOTTSGMQ
TTGEBVAAATGGRATPAGAGATVAGSARSANVAASVTSATASAMGVDDSSAGVNSTRA
MGRQNSNGENSVAGLYNSFVRQSGANVAVALGANAGASGAVSFGSSRYEANYVS
VQSGNGRGCPATRTITNYSAGALADASDANNGCFQFSLNITSFISGGKALADQGVN
FPAPYVLOGASYNVNGAALTALDSKVELDARGTAAATATATPYSRTAPVPAVVAS
ATVASSDVTGDSASVOGPTPAVYVGTATPATSTPRTAVGNAANHTTGATIGGASVANG
PBDTIGSNARVNMADSGTAVGANQDIAAVNANVAMGAGVSAASCTAIGQAKRASA
QCAVALVQSAADRANTYSVSGSGCERGVAVANVAGATRTIDMANGQDLSGVAAANSTG
DSRISAMDDSEFTTQGDIEDRLRQNRNTRLDQKAGMSALMMSASVGLSASQNTIGAG
VGFQNGESLISGVQRAISPATVYIGGALSIGDSDSSISGVAGFGW"
8306..10186
/gene="xCC0659"
8306..10186
/gene="xCC0659"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="AA039975.1"
/db_xref="GI:21111665"
/translation="WIDKNSRLNPIETGAIIIMALGASGTVAPADALQKEPTQAAPAA
SDFNSLIVRKDKGTAAASDBRSSKLGIVQVQSVAGRAAPYTGARAQATAKATYLEKGI

CDS

gene

BASE COUNT	1792 a	3332 c	3541 g	1610 t	
ORIGIN					RO"
Alignment Scores:					
Pred. No.:	31.4	156.00	Length:	10275	
Score:	38.77%	Conservative:	80	146	
Percent Similarity:	25.04%	Mismatches:	247		
Best Local Similarity:	6.47%	Indels:	111		
Query Match:	1	Gaps:	25		
DB:					
US-09-825-414-7 (1-486) x AE012164 (1-10275)					
0y	5	ArgArgValGInGInProProValThra1aThraSpsrPheArgThr-AlaSerAsp--	23		
Db	5870	CAGCGAATCCAGCGCGCCCTGTCTCGCGACCTGGCGCTACCGCCAGCGGCGTCCACGGA	5929		
0y	24	-----AlaSerLeuAlaSerSerSer-----	30		
Db	5930	CTACAGCACCCGACGTGGTGGCGCGCTGCTGCTGCTGCGCGCCACGCGGCGTGGCGCGTGGG	5969		
0y	31	---ValaArgSerValSerSerAspGIn1aArgGIn1aLeaSn-----	43		
Db	5990	TGAATTACAGCAGACCGCGCGAGTAAGAACGCGGTGGCGGCGGACAGACGTTCTTCGG	6049		
0y	44	-----	44		
Db	6030	TTTCATCCCGGACGCTGGCGGCGACCGCGTGACAGCGCCTTCGGGTGGCGGTGCATGGCC	6109		
0y	44	AlleAlaAspTryLeuThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSe	64		
Db	6110	CACGGCCGATTACACACCGCGATGGCGTGGAAATTCATGCGCAGCGCGCTCACAGCGCAG	6169		
0y	64	rAlaAspGlyGInAlaAlaValaAspVal-----HisAsnAl	76		
Db	6170	TGCATTGGGCGCAGACGCGCGCGCATTCGCGGATTAACGCGGTGGCGATCGGCGGCAACAG	6229		
0y	76	aglInIleThrAla-LeuIleGluThrArgAlaSer-----ArgLeuHisPheGluGlyg	94		
Db	6230	CCGCCCGCATGGATGGTGGTGCAGCGGTGGTGGCGGTGGATGCGTGGCCACCGGCATCA	6289		
0y	94	luThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaT	114		
Db	6290	CAGCAGCCGGCGCGCGCCCGCAGTCAACGTAATGGCGAANAACCGCGTCTCGTGCGCTA	6349		
0y	114	hrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnIleTyrM	134		
Db	6350	CACATCGTTCGGCGCG--AAGCGCGGTAAACGCTGTGCTCGGTGGTGCATTCGCGG	6406		
0y	134	eGInProAlaAlaLeuAsnLysGlyAspTryPleuProAlaPro-----LeuLysProLeuT	152		
Db	6407	TGCCACCGCGCGCGCATTCGGTGGCATTTGGGCTCGCGGCTCGCGCATTAACGAAGCAAC	6465		
0y	152	hrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGInValGlyThrIlysmeteta	172		
Db	6466	CGGTGTCGCGT---GGTGGCGGCAATGGCCCGCGCGCGCGGCAACGCGTGCATGCTCA	6522		
0y	172	sPArgAlaThrGlyAspLeuHisTyrIleuSerAlaSerProAspArgLeuHisAspAlam	192		
Db	6523	ATGTACGGCAGCGGCAATCCGCCACCGATGCGGTCACAAAGGCGCAGCTG---GATGAT	6579		
0y	192	eValaAlaSerValLysArgHisSer-----ProSerLeuAlaA	205		

Db	6580	TGGCGCGCGGATGTCAGACACACACAGCGCGATGTGTACACACACACCGCGGACGGCGGTGGCC	6639
Oy	205	rglInValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgT	225
Db	6640	GTGCACACGGGGATGGCGCCACCGCGCGGGTGGCGGATGGCCACCGCCAGCGTGACGCA	6699
Oy	225	hrValLeuAlaProAlaLeuAlaSerAlaArgProAlaValGlnGlyValAlaAspLeuGly	245
Db	6700	GGGTG--GGCGTACCCCGACGGACACTGCTGCGCCACCGCGCCAGCGGATGGGTG	6756
Oy	245	aI---SerMetAlaGlyGlyLeuAlaAlaAsnAla-----G	256
Db	6757	TGGACAGTACGGCCACGGCGCGTGAACACACACCGCATGGGGCGGACACCACTCATTTG	6816
Oy	256	lyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGly--GlyAlaL	275
Db	6817	GGGACACGGCGTTGGCGGTTCACACTCTTTGTGCGCCAGACGAGCGGCCACACGGCG	6876
Oy	275	euValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerCulLysAsnAspTrpL	295
Db	6877	TGCGCCCTGGGTGCCAATGCGGGTGGCCACGGCGCGCATCGGTGGCGGTGGCTCCGGTT	6936
Oy	295	eugLIAlaTyrLysAla-----IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnA	313
Db	6937	CGCGCACTATGAAACCAACGCGTGTGGTGCGGACGGCACGGCGCGGTGGCCCCG	6996
Oy	313	IaGlyLysArgMet-----AlaGlyLeuProLeuAspMetAlaThrAspAlaM	329
Db	6997	CGACGCTGCATCGTCATAGTGGGGTGGCGGGCGATTTGCCGATTCACGACCGATGGGA	7056
Oy	329	etGlyAla-----ValArgSerLeuValSerAlaSerSerLeuThrGlnAsnGlyL	346
Db	7057	TCCACGGCGGCAGTTCCTTCACAGTCGCACAGAAATACCGCCAGCTCCCTCGCGGGTGG	7116
Oy	346	euAlaLeuAla-----GlyGlyPheAlaGlyValGlyLysLeuGlnGlnMetAlaTrpL	364
Db	7117	CGCGCATCGCGCGCGCAAGGTGTGTTCTGTGGCGCCGCACTACGTCTACCAAGGTGCACGT	7176
Oy	364	ysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeu-----	381
Db	7177	ACAAACAAGCGGTGCGCATATGACCGCGTGGACACAAAGGTACACGCGTGGATGCAC	7236
Oy	382	--AlaGlySerAlaAlaValPheAlaGlyTrpThr-----ThrAlaAlaLeuT	397
Db	7237	GTGGCGGACCGCAACGGCTACCGCTGCACGACGCGTTTGGCTGGCACTGACGCGGTT-	7295
Oy	397	hrThrAspProAlaValLysLys-----AlaGluSerPheIleGlnAspTrpY	413
Db	7296	-----CGCGGGTGGCTGCATCCGCGACGGTGGCAAGCAGCATGTCACTGACACA	7347
Oy	413	alLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaL	433
Db	7348	GGCGAAGGTGAGGGACGCGACCGCGGTGGCGGTGGTGGC--ACGGCACACCGCGGGG	7404
Oy	433	ysThrValLysAspMetGlyGlyGlnAlaIleThrHis--ThrGlyAlaSerLeuArgAsn	452
Db	7405	CGACTCTGCACGGCGTTTGGCAACCGCGCTGTGGCCACACATCTACCGGACCGCGATTG	7464
Oy	453	ThrValAsnAsnLeuArgGlnArgProAla-----ArgGlnAlaAspIleGlnLeuGly	470
Db	7465	GTGGCAGCGCGGTATCGCGATGGCCCCAACGACACCGCATTTGGCAGCAATGACCGGTCA	7524
Oy	471	glyThrAlaAlaSerProSerCulLeuPro-----PheArgProMet	484
Db	7525	ATGCGAGTGGCAGCACCGCGGTGGTGGCAACACGCAATTTGCTGGGTGGCGACCATATG	7584
Oy	485	Arg	485
Db	7585	CGG	7587
RESULT 9			
AE000234			
LOCUS	AE000234	13431 bp	DNA linear BCT 01-DEC-2000

```

DEFINITION Escherichia coli K12 MG1655 section 124 of 400 of the complete
genome.
ACCESSION AE000234 U00096
VERSION AE000234.1 GI:1787633
KEYWORDS
SOURCE
ORGANISM
Escherichia coli K12.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 13431)
Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
Blythe,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
Gregor,J.J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
Mau,B. and Shao,Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
TITLE
JOURNAL
MEDLINE
97426517
PUBMED
9278503
REFERENCE
2 (bases 1 to 13431)
Blattner,F.R.
Direct Submission
JOURNAL
TITLE
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
3 (bases 1 to 13431)
Blattner,F.R.
Direct Submission
JOURNAL
TITLE
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
4 (bases 1 to 13431)
Plunkett,G. III.
Direct Submission
JOURNAL
TITLE
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amher.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ***The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
Source
location/qualifiers
1..13431
/organism="Escherichia coli K12"
/strain="K12"
/db_strain="MG1655"
/sub_xref="taxon:83333"
<1..1145
/note="IS5"
complement(97..1077)
gene

```

```
/gene="trrs_5"
/note="b1370"
complement(97..1077)
/gene="trrs_5"
/function="IS, phage, Tn: Transposon-related functions"
/note="f136: 97 pct identical to Y152_ECOLI SW: P03837 but
has 12 fewer N-terminal amino acids"
/codon_start=1
/transl_table=11
/product="IS transposase"
/protein_id="AAC74452.1"
/db_xref="GI:1787634"
/translation="MSHQLTFADSEFSTKRQRTKEIFELSRMEQILPMQNTAVIEPE
YPRKNGRRPYPLETMLRIHQMOHWNLSDGAMEDALYEIASMRLFLSDSLPDR
TTIMFRHLLEOHLAROLFKTIRWMLASGVMVTOGLVDATIEAPSSKNQEOOR
DPEHQTCKGNOMHEFMKAHIGVAPKSLGTHSLVTTANEDLNQNLHGEEOPFS
ADAGOGAPORELAEVDVMLIAPRGKTKTLKONPRKKTAINITYMKASIRARVE
HPRRIKRGPFYKARKGLKNDNQLAMLEPTLANLFRVDMIRQWERSQ"
709..736
/note="factor Sigma70; predicted +1 start at 1426416"
729..758
/note="factor Sigma70; predicted +1 start at 1426438"
874..1335
/gene="b1371"
874..1335
/gene="b1371"
/function="orf; Unknown"
/note="o153: This 153 aa ORF is 57 pct identical (0 gaps)
to 66 residues of an approx. 208 aa protein V10M_LAMB SW:
P03701"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74453.1"
/db_xref="GI:1787635"
/translation="MLHAVNTHGLORIGPSATARGIKRUDSGHTLPMQNLHAGE
ENLESGTALSAETIGEGELMADVPLGCAPDEYDILSGTCSHLPKGTHVLTGS
DDGRHNTSLMAGVQFNFTESVAIDIAIEGPGSGDWRTDGFVGVYKF"
complement(1230..1256)
/note="factor Sigma70; predicted +1 start at 1426896"
1304..1330
/gene="b1371"
/note="factor Sigma70; predicted +1 start at 1427010"
complement(11346..11363)
/note="central position to predicted promoter: -115"
/bound_moiety="phob predicted site"
1394..4762
/gene="b1372"
1394..4762
/gene="b1372"
/function="putative membrane; Not classified"
/note="o1122: This 1122 aa ORF is 76 pct identical (1
gaps) to 381 residues of an approx. 408 aa protein
Y401_LAMB SW: P03764"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC74454.1"
/db_xref="GI:1787636"
/translation="MNMVAKISGLVKGDTGKRPVONCTIOLKAKRNTVVVNLASPN
PDAGRSKMEYEGQYVYIIVLEGFPFSSAGTTIVYDSQPGTINDPLGAMTEDDAP
EALRRFLMVEVARNSAVAOHTAAKKSADSDSTSRARATHAADAUSARASISQ
AGQASASQASAGASTATKATASAAAESEKSAATSAATSAATSTNASSISQ
SAATSAATATTKASEATSAADAASKEAKSEPTNASSASSASATAGNSAKA
KTSPTNARSETTAGOGASAAAGSKTAASASASASTSAGQASASATAAKRSASAS
SASATTKACEATPEOASAAARSASATSTNNKASTSESSSTTAASSASSASSA
SSASAKDEATROASAKSSATTAATSTATPAGSATAAOSKSTAESAAATRAETAAK
AEDIASAVLEDAESTTKRGIVQSSATNSSETLAAIPKVKSAFYDAERLAKDQNG
ADIPDKCCFLTNINAAVSKTDFADKRCGRVYRVNAPATGKXYPVVVMASVSSEL
ASRYITLTARTAGDPNNCEFNQFVPGWTDGGRAYGMFOYONNEKAHISIMMS
NKGDRLSVFYVGAAPVFAETIEDGISAPADADIVYNTQYTKFGATNTATNCIAD
VILDFKSGRGFTESHSLIVNDNISCKKLPFTDELVAAGGQIIRIGGETYCALNRDAS
KYILLTNGDVTYGMWTLRPFALDNTGELVIGTKLSASLNGALNTATATKIQTPRRVS
```

```
GVFEDGSKDITLTAAVAAAFARRATDTYADADGVPMNAESGAVNTRGDSYIIVNF
YTGVCSCRTLOMKAHYRNGGLFYRRSSRDYGFEDEAAEYTSKPLPESYPVGAPIPM
PBDTPSVGVALMOGOAFDPSKAYPLAAVPSGVIPPMRGWITIGKPSAGRAVLSQOD
GIRKSHNSAASSTFDGCTKTSSEPDQCTSTNTGATHTHSVSSTSGASAGHTSLANV
NTASANGASASATRLSYVHNQYATSSACAHHTSHLSGTPAASAGAAHYVGICAHHTS
VAISGHGHTTVVAAAGNAENTVKNIAFNITVRLA"
4762..5337
/gene="ynaC"
/note="b1373"
4762..5337
/gene="ynaC"
/function="orf; Unknown"
/note="o191: This 191 aa ORF is 94 pct identical (0 gaps)
to 191 residues of an approx. 200 aa protein Y194_LAMB
SW: P03740"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74455.1"
/db_xref="GI:1787637"
/translation="MAFRMSEQPTRIKIYNLAGTNEFTIGEGDAYIPPHGLPANSTD
IAPDIPAGVAVFVNSDESSMHLVEDRGVTVDVASGNALFISELGPLPENTWTISP
EGSEFQKNGVAVAKDDEAEKLEIRAEETKKNLMQVASHIAPLDADADLEATEE
ISLEAMKRYRVLLNRYDSTQADIEPALP"
complement(5435..6025)
/gene="b1374"
complement(5435..6025)
/gene="b1374"
complement(5435..6025)
/gene="b1374"
/function="IS, phage, Tn: Not classified"
/note="f196: This 196 aa ORF is 76 pct identical (0 gaps)
to 193 residues of an approx. 200 aa protein TNPO_ECOLI
SW: P05823"
/codon_start=1
/transl_table=11
/product="putative transposon resolvase"
/protein_id="AAC74456.1"
/db_xref="GI:1787638"
/translation="MSRITFVCRISITLIDOTTENQREIESAGFKIRPOOIEEHISGS
AATSERGFNRLRLRLUCGDQILVTIKDRIGCANMIDIRVQDLTGTGRVHCLAGG
IDLTPGKMMQOVISAVAEFERDLLEFRHSIVIRAGAKRFRGPRPVLNEBQKQVY
FERIKGSVISATIAREFKTSRQITLRAKKLQPRD"
complement(6033..6061)
/note="factor Sigma70; predicted +1 start at 1431699"
complement(6342..6608)
/gene="ynaE"
/note="b1375"
complement(6342..6608)
/gene="ynaE"
/function="orf; Unknown"
/note="f186: This 88 aa ORF is 32 pct identical (4 gaps)
to 71 residues of an approx. 1112 aa protein STV_YEAST SW:
P07806"
/codon_start=1
```

Alignment Scores:

Pred. No.:	Length:
Score:	13431
Percent Similarity:	155.50
Best Local Similarity:	35.49%
Matches:	121
Mismatches:	60
Indels:	220
Gaps:	109

DB: 1 21

US-09-825-414-7 (1-486) x AE000234 (1-13431)

Qy 19 ArgthrlaseraralaserleualserSerSerValargSerValserSerAspGln 38

Db 1742 CGTAACCGGTCCGGGTGGCCAGAACACGGCCGGAAGATGACGACGAGGAGCC 1801

Qy 39 GlnArgGluileanalailealaasptyltleuThrAspHisValpHealaalanlelyls 58

Db 1802 AGCACATCAGCCCGGTGAGCGGCAACCATGCGGCTGAT-----GCTGGCGATCTA 1852

Qy 59 leuProfoalaasSeraralaaSpGlnlaalaValaspValHisasnAlaGlnIle 78

```

Db      1853 GCACGGCCAGCCAGCAGCTCAGCCGAGCCGCGCTCGCTGAGCTGAGTGCAGGAAAGGGGATG 1912
Qy      79   ThrAlaLeuLeuIleuThrArgLysArgLeuHisPheLeuIleuLysIleuThrProAlaThr 98
Db      1913 AGGGCA-----GGAAAGGATCAAAAGAGCCACT 1942
Qy      99   IleAlaAsp---ThrPheAlaLysAlaGlyLysLeuAspArgLeuAlaThrThrThrSer 117
Db      1943 GAAGCATCAAAAGAGCTGCTGCGCTGCAGAGTCTCAAAAAGCGCGCGGACACAGTCC 2002
Qy      118 GlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnIleuMetGlnProAla 137
Db      2003 GGTGGCGGCAAAAGCTCAGAAAGCAATGCTTACAGCTCAATCAATCAGACAGCAATCT 2062
Qy      138 IleAsnLysLysPhePheProAlaProLeuLysProLeuThrProLeuIleSerGly 157
Db      2063 GCATCCAGCCGCG-----ACCAAGAGGATCAGAA 2092
Qy      158 AlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAsp 177
Db      2093 GCTGGACATCGGCGCGGATGCGGCGGCTCAAAAGAGCGGCAAAATGATCA---GAA 2149
Qy      178 LeuHisThrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAla-----Ser 195
Db      2150 ACGAAGCATATCAAGCGCGCACTAGTGCAGTCTTCGCGCAAGCGGCGGAGGAAATTC 2209
Qy      196 ValLys-----ArgHisSerProSerLeuAlaArgGln 206
Db      2210 GCGAAGCGGCAAAAGCTCCGAGACAGAACGCCAGGCTTCTTGAGAACGCGGAGCAG 2269
Qy      207 ValLeuAspThrGlyValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArg 224
Db      2270 AGGGCTCGGCTCGGCGAGGCTCAAAAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2329
Qy      225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly 244
Db      2330 ACAAGTCCGCGGCGAGCTCTCAGCAGCTGCCAGCGCGCGGCAAAATCGCAGAAACGCC 2389
Qy      245 ValSerMetValGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 254
Db      2390 GCATGCTGCTCTCAACACAGCCACAGAGGCTGGCGAAGCCACTGAAACAGCCACGCA 2449
Qy      265 GlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeuLysAspLysGluPro 284
Db      2450 GCAGCGAGGCT-----GCTTCCGACGAAAGATCCGAAACG 2488
Qy      285 LysAlaGlnLeuSerGluGlnAsnAspTrpLeuGlnAlaTrpLysAlaIleLysSerAla 304
Db      2489 AACGCCAAAGCGTCGGAACACAGC-----GCAGAAATCTCAAAAGCGCTGCCGATCG 2542
Qy      305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaLeuProLeuAspMet 324
Db      2543 TCAGCCAGTTCGGCGGCGTCAATCGCATCATCG-----GCGTCTCTTCAAAAGTAG 2596
Qy      325 AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn 344
Db      2597 GCGACCAAGACAGCGTCAGAGAGAGAGC-----ACGCCACAGACGCGATCCAGAAC 2650
Qy      345 GlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnLeuMetAlaThrLys 364
Db      2651 GCGACAGAGCGTCTGCGCAGT----- 2671
Qy      365 AsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySer 384
Db      2672 -----GCGACGGCGCAGCTCAGAGCAAAAGTAG-----GCGGAATCC 2710
Qy      385 AlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrThrAspProAlaValLysLys 404
Db      2711 GCGGCAAGCGCGC-----GAGACAGACGCTAAACGG 2743
Qy      405 AlaGlnSerPhe-----IleGlnAsp-----ThrValLys----- 414

```

```

Db      2744 GCAGAGATATTCATCCCGCTGCGCTTGAGAGATGCAAGTACAGGAAAGGGGATG 2803
Qy      415 -----SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu 431
Db      2804 GTACACCTGACAGCTGCGACACAGTACGTGTGAAACCTGCGGCAAGCGGCGGTGAT 2863
Qy      432 AlaLysThrVal-----LysAspMetGlyGlyAla 442
Db      2864 GTAAATTCAGCGCTATGACATGCAGAGAAAGCTCTGCAGAAAGACCAAGAACGGCGGTGAT 2923
Qy      443 IleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArg----- 458
Db      2924 ATACCCGATTAAGGAGTCTCTCTG---AACACATTTAAGCGCGGTCACTAAACAGACTTT 2980
Qy      459 -----GlnArgProAlaArgGlnAlaAspIleGluGluGlyCysIleThrAlaAspPro 476
Db      2981 GGTGATTAAGCGGTATGCGTTATGTGCGGTTAAACGCTCTGCAGTGCACATCTGGA 3040
Qy      477 SerGluIleProPheArgProMetArgSer 486
Db      3041 AAATATTACCTGTCTGTGTATGCGTCT 3070

RESULT 10
AE007019      15424 bp      DNA      linear      BCT 27-APR-2001
LOCUS        Mycobacterium tuberculosis CDC1551, section 105 of the
DEFINITION   complete genome.
ACCESSION    AE007019 AE00516
VERSION      AE007019.1 GI:13881105
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE
1 (bases 1 to 15424)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J.F., Deboy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,M.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouli,H.,
Gill,J., Mikula,A. and Bishai,W.
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL
Unpublished
2 (bases 1 to 15424)
Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
Peterson,J., Deboy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E.,
Kolony,J.F., Nelson,M.C., Umayam,L.A., Ermolaeva,M.,
Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khouli,H.,
Gill,J., Mikula,A. and Bishai,W.
Direct Submission
TITLE
Submitted (25-APR-2001) The Institute for Genomic Research, 9712
JOURNAL
Medical Center Dr. Rockville, MD 20850, USA
FEATURES
source
1..15424
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="Clinical strain"
complement(171..581)
/gene="MT1491"
complement(171..581)
/gene="MT1491"
/note="identified by Glimmer2; putative"
/codon_start=11
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45754.1"
/db_xref="GI:13881106"
/translation="MTWADSGRAPYRRMKTLTOALNADKTRVEOVYLGICK
TWAEINSSLSQINSTEYERLEBGDLHEBGLTSLDLDLARKLIVLVEPVAIVDRIDITY
SLGETVMSPLSTEHAVRGVLDRLRNRTVHEPTN"
complement(598..1341)
gene

```

```
gene
CDS
/gene="MT1492"
/complement(598..1341)
/note="similar to GB:U00013 PID:466869; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glucosamine-6-phosphate isomerase"
/protein_id="AAK45755.1"
/db_xref="GI:13881107"
/translation="MSSSIEIFPPDSILVAAAGKRLVGAIGAAVAARGOALITLVGG
NGIALRTLSAQAQIEMSKVHLFWGDERVYEDDEDERITACARALLNHNIDIPNOY
HPMASSDSDPGSDDAALALAEQVLAASAARQDPAPNDVHLGSGPGHINSLEPHS
PAVLESTKAVVAVDSPKPPRRITLTPAIQORSREWVLVSGPKADAVAAAGGAD
PVYPAAGAVGQNTLWLDRLDAAKLRPS"
/complement(1338..2249)
/gene="MT1493"
/complement(1338..2249)
/note="MT1493"
/note="similar to GP:5459405; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxopycic protein Opca"
/protein_id="AAK45756.1"
/db_xref="GI:13881108"
/translation="MIVLDPDTTAVNKKLDLEKIGAVAMGFVLTLLIAPDSEAM
LEESIEAANDSHEHPSRIIVMRGDPADRLDQALRGADAGCFVYLRISGLP
AGHADVLPFLPDIPYVAMPPIADAPADODLGLAIRITDATNADPLSILKS
RLAGYAGDGLDAMSRITVYWRALLTSADDPHEPISAVSGIKTEPALDYLQWLA
SRIBGPVRAAGELKVELYRNSRTIVLSRPOGITAATLTRGKPAALVPLARVYGECL
LAEDLRDPDGLTICALEGIKKVOYR"
/complement(2302..3846)
/gene="MT1494"
/complement(2302..3846)
/note="MT1494"
/note="similar to GP:5459404; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate 1-dehydrogenase"
/protein_id="AAK45757.1"
/db_xref="GI:13881109"
/translation="MKPAHAASWRNPLRDKRKRPLRIAGPCGVIFGVYGLDARKK
VMPAYVDLANRGLPPTSLVGFARMDSTDFGVVYNAVQEHCRTPPCQOMRRLA
EGFRVPTGDDDDPAQLAETLEKIDAEKRGHAEFLAIPKSPVYCEQLHKSQ
LARPGDRMSRVYIEKPRGHDLASARELKNAYAPPEAVFRIDHYLAKETVONILA
LRFANOLFDPIMAHYVDHYOTTAEDIGLGRAGXYDGIAGARVYONHMLQDLALT
AMEPVSFHPALQAEKTKVLSATRLAERLDOTTSRGQIAGWQGEKVGILLDEGF
AEDSTTEFAATLEVDTRRMAGVPFYLRIGRLGRVTEIALVFRRAPLHPEFDTMT
DELGTNAVIRVQDDEGVTLRFGSKVPGTAMERVDVNDFSYGSFAEDSPAYERLI
LDVLLGEPSLFPVNAEVLWEIIDLPALEHMAAGCTPDAYEAGTWGPESSLEMLRRTG
REMRP"
/complement(3843..4964)
/gene="MT1495"
/complement(3843..4964)
/note="MT1495"
/note="similar to GP:2398718; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transaldolase"
/protein_id="AAK45758.1"
/db_xref="GI:13881110"
/translation="MTAQNPNLAALSAAGSVWLDLSRDLRSGNIOLITDSYVG
VTNPSITOKALSEGHTYDAQIAELAAAGADYATIRYTTDDVASACDVLVPQWEDS
DGVDGRVSIEDVPLAHTEKTIQOAILKMLKIVDPNLFIKIPAKADLPALISAVLAE
GISVNVTLISVQRYREVDAYLTGEMKARQAGHLSKIHSAFFSVRSRVTEIDRL
DRISGRQALELGEQAGVANARLAYAAAREVEDSDRSLKVDGARVQORPLMASTGVK
NPDSYDTLVTELVAPHTVNTMPEKTIIDAVADHGVIOGDVTGTASDAQAVFDQGAIT
GIDLTDFAVILEEGVRKEEASWNNLLOETRAHLDITAAQ"
/complement(4981..7158)
/gene="MT1496"
gene
```

```
gene
CDS
/complement(4981..7158)
/gene="MT1496"
/note="similar to GB:U00013 SP:P46708 PID:466894
PID:2398717; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transketolase"
/protein_id="AAK45759.1"
/db_xref="GI:13881111"
/translation="MRHCKRDSRTRYVTDHLDASRTVEVTTLEISALTRPRHDDM
TEISPAADTIRIVADAVQVGNHPTASLAPLVATLTFORMRHDPSTHYLGRD
RPVLSAGHSSTLYLTOLYGGFGLSESLRPMWSTKTPRPERHPPRGTITGTPG
GOGLASAVGAMASRYERGLDPPRAEPASFPDYITYIASDGIIEBYSSEASLLA
VQQLGNLIVETRNQISTEDTNTALCELDIARTARAEMHVEVEGGEENVGIEALA
NAQAVTDRPSFIALRTVGYPAFMLMDGKAHGAIDDEVAARKIYGPDPDFTFVY
REDVLTHTRGVLVARGKQAHERMQLFEDMWARREERALLDRLAOKLPDMDADLP
WEPGSKALATRAASGAVLSALGKPIELMAGSADLAGSNNTTIKAGDSPSTKE
YTAHYHGTILHGVREHMGAILSGIYHGTFRVYGGFIQFSYMPRAVLAALMDI
DTIYVHTDSITGLGEDGPTTHPIELSLAKRILRISYVRPRDANETAYMTILIARRN
GSGPVGLILTRQGVPLDGTDAEGVARGVYLSQGGLOPEEDVAILIATGEVOLA
VAAOTLLADNDILARVYVSMPCLEWFEAPYPRVAVLPPVYSAVVAEAGAACMWHOL
VGDIGEIVSIEHYGESADHKTLFREYGTABAVAAERALDN"
/complement(7521..11747)
/gene="MT1497.1"
/complement(7521..11747)
/note="MT1497.1"
/note="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="PE_PERS family protein"
/protein_id="AAK45760.1"
/db_xref="GI:13881112"
/translation="MMSLVIAPETVAAALDVARIIGSSIGANANAAGSTTVLAAG
ADEVSAIATFLFEGSHAREYQAIISTOVAAHFDRPQOTLSAGVYVAEATNAALATL
EHNVTLMANAPTOALIGRPLIGDCAAGAPGTCQAGGAILMWSGAGSGAPROVCG
AGGAGLGTGTGAGAGAGAGAXAGGSGGGVLDNGVYGAGGSGSLGARTGAGGNA
GLFVGGTGXGPGGPGPGVGTGAGGLGTLVAGHGGAGGAPGLPGVGGHGVG
GAAGLIGVGGHGGAGGJGAEVAGAGBDLSLPHGTSGVGGDADGDTGGVGGWLAG
GGAGAGGVGTGGAGAGGAFRRALIVADNCGDNGGNGGAGGAGGAGAGGLISL
GGGAGGAGGTGGAGGAGVGGDGGAGGPGQANAGAGAGGAGGPGAGGAGCTGAGS
TTGAAGAAGASPTSGNGAGAGCNANHPSSGKAGAGNGAGAGGLVNGGAGAGCNC
APGAPSGGDPNNGGGGAGAGGKGGDDGAQGGAGGAGGAGGNGVGTATGLM
GLGAGADDTDGGKNGGAGGAGGAGGAGGAGALATYHODGSMGAGAGNGAGAGTGG
DGGNGAKTPTFNGGNGVAGGNGGNGSRLIGAGGIGGAGSTAGADGARATPTSGSNG
GTGGGANATYVAGGAGSGAGGNGGLVNGGAGGAGGAGGAGVAGSPTTAGSGSNG
GONGAGAGAGGAGGDPDGGGTGAGGADGAGGAGGAGGAGNANAMATTPAKGDDG
GGGERAOGNGNGGSGPGLAANLIRGONTGVYSGSGKAGAGGLAAGDGNKANRPAE
GDNGHGNGNGNPGAGQGGSGGAGSTPGAKGAGHFTPTSGDGDGNGNGNSOVVG
GNGDGGNGNGGSGAGTGTGNGRGDDGAFGGMASNATNPGENGNGNGNGGNGAGGAG
GAGLNGGNGGAGNGGLGFGGNGAAGNAYVAGAPGPGAGGHHGNGGNGAGGAG
GOGVYSDAGAGGAGGDDGAPGGANGNGNGAGAFAGGGGNGGDDGNGNGAGAG
PGGTGSTGKAGAPGASTYHDDGNGNGHGGHAGSANGNGPGHGGNGGNGGTGANGNG
GITGTGAGSGTGAAGVLCITNCGDGDGDRGNGRNGRCGCLTGAGGNGSGTCTPRN
GTGGNGAGSGLVITSFGDGGGGRGDDGRTGCGDGLGSSSPGGTPODMGTGTGGTG
GTGGGANGAGJLTGRGTGTGNGNGNCTGAGAGTGGTGNGSGQPGMGNGAGGFGG
GNGFVAGGRCMGSGGCTGTGAGPPTGTGTGCGHGGGCGGCGGCPSTILGLGLGG
LGSPISTGTGAGGAGGCGGFGGCGGFEV"
/complement(11938..12096)
/gene="MT1497.2"
/complement(11938..12096)
/note="MT1497.2"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45761.1"
/db_xref="GI:13881113"
/translation="MVDRVTVAAPQGRFTRLRHHTIGTARQPEPRSGOFSHQDENRY
IGLISYVP"
/complement(12146..13072)
/gene="MT1498"
/complement(12146..13072)
gene
```


AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.

TITLE The systematic sequencing of the *Escherichia coli* genome in Japan

REFERENCE 3. (bases 1 to 16902)

JOURNAL Unpublished

AUTHORS Mori, H.

TITLE Direct Submission

REFERENCE Submitted (14-DEC-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan

JOURNAL (E-mail:hmorigtc.aist-nara.ac.jp, Tel.81-7437-2-5660, Fax:81-7437-2-5669)

COMMENT Collaboration Information:

Project: The Japan E.coli genome DNA sequencing project

Group: The Japan E.coli genome DNA sequencing group

Members: (1995.4 - 1996.3)

Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.

Headed by:

Name: Takashi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishori@nibb.ac.jp

Information operator:

Name: Hirotsada Mori

Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan

E-mail: hmorigtc.aist-nara.ac.jp

URL: The Japan E. coli genome database

http://bsw3.aist-nara.ac.jp.

location/Qualifiers

1. 16902

/organism="Escherichia coli"

/strain="K12"

/db_xref="taxon:562"

/map="30.7 min"

/clone="Kohara clone #264"

/clone_1db="Kohara lambda miniset library"

/note="Nucleotide position 1428466-1445367 from the initiation site of *Pha* (0 min.). This clone is from Kohara lambda miniset library"

813. 1841

/note="ORF_ID:0263#17

similar to [SwissProt Accession Number P03736]"

/codon_start=1

/transl_table=11

/product="Minor tail protein precursor H."

/protein_id="BA14970.1"

/db_xref="GI:1742238"

/translation="MAEOTSRLAIIIDTGAKNNADMLTSLVMTQAGTEAANSACK VTKAEDELKALAKAIDPVGAIDVGRARYSELKPFDKGLIDKEEYELVRIKLN ETTEELSEGVAAQARBEAKGKLAQAQOAFORMLDKIDPLAALRLNLEQOHDEL MAAPASKINGSPENTSRKIQTRERLTEBAQERPAKAHDEQVALDRLAQDDP VGTAFNLEVEQKOLNEAKAKGMLSPMEYELSGKLRAHSELEVTQSLSKGTGMSAK QTAFAFMLPAQMTDVIYVSTGSPMVMQOGSHSAGMETRIASLPVTAIPAGMKI PCVTAGSA"

1817. 1972

/note="ORF_ID:0263#18

similar to [SwissProt Accession Number P03701]"

/codon_start=1

CDS

gene

/transl_table=11

/product="Outer membrane protein Lom precursor (ORF 206A)"

/protein_id="BA14971.1"

/db_xref="GI:1742239"

/translation="MNRNMFSSVAVGFSVRVNEFSAVAMAGNAVSRVSTSGDYLKVT DNRGRE"

complement(2105. .3115)

/gene="1S5"

complement(2105. .3115)

/gene="1S5"

/note="ORF_ID:0263#19

similar to [SwissProt Accession Number P03837]

start codon is not identified yet"

/codon_start=1

/transl_table=11

/protein_id="BA14972.1"

/db_xref="GI:1742240"

/translation="FIRKASQRDIMSHQTFADSEFSTKRQRTREKIFELSRMQLPWP QNMTAVLEPPYPRKNGRRREPLETMRIRHQMHWNLSDGAMEDALYETASRLFLAR LSLDSALPDRITLIMNFRLLEQHLAQLKRTINRMLEAGVMTQGLVDATIEAP SSTKRNKQORDPMMHQTCKGOMHFGKKAHIGVDASGLTSHLYTTAANDHDLQGN LIGGEQFVSADAGYOGAPQRELEAVDNDMLAERPGKYTKLQNPRRKKTALINLEY MKASIRARVHPRIIRKRFQFVKARKIKGLKNDNLAMLTLANLFRVDMQROMER SQ"

2441. 2809

/note="ORF_ID:0263#20

similar to [SwissProt Accession Number P03838]"

/codon_start=1

/transl_table=11

/protein_id="BA14973.1"

/db_xref="GI:1742241"

/translation="MLISMKQITOLIEVLMVGRGQDVGQATIGIDTNYGLHAKYPL IAFGLMHRITALLPYLGRAGCLNDSGHIQSLAGHNDAFCQPAIDGLQLGLQGLML LEQVAEIHDDGALRQGAHQG"

2882. 3343

/note="ORF_ID:0263#21

similar to [SwissProt Accession Number P03701]"

/codon_start=1

/transl_table=11

/product="Outer membrane protein Lom precursor (ORF 206A)"

/protein_id="BA14974.1"

/db_xref="GI:1742242"

/translation="MLHAVNTQGLQIRGSAIARLGIKRLDDSHILPQNLHAGE ENFSGFGLTASAEFTIGEGELMAHDVLCGAPPEYDLISGTSHLPKGTHTPVLTGS DQGRHSTSLAMWAGVQFNFTESVAIDIAEGRGSGDMWRDGRITVGGYKF"

3408. 6770

/gene="37"

/gene="37"

3408. 6770

/gene="37"

/note="ORF_ID:0263#22

similar to [SwissProt Accession Number P03744]"

/codon_start=1

/transl_table=11

/product="Tail fiber protein GP37."

/protein_id="BA14975.1"

/db_xref="GI:1742243"

/translation="MAVKISGVLDKGRKPVNCTIOLKAKRNSTVVTNLTASEND EAGRTSMDEVGQYVILLYEGFPPSPNAGITTYEDSQPRLTDLFGAMVEDDARPA LRRELWVEEVARNASAVAOINTAAKASADASTSABEATFHADADSKRAASTAG QAASSAQSASSAGTASTKATEASKSAAAESSKAAATAGAKTSSETVASLQSA ATASATATTKASAPATISARDAASKEAKASSETNASSASSAASATTAAGNSAKAAT SETNARSSETTAAGSAAAGSVTAAASASASTSAGASASATTAAGGAESAASAS SATTATGATGEATBOASAAARSASAKSETNKAASSETNASSAGSAAASASSAS ASASKDATQASAKSATTATSTKATEAAGSATTAAGSSTESATTRATKARAE DIASVALEADASTTKGIVOLSSATNSTSETLATPRAVKSATDNKRLQKQONGAD IIPKGEFLNNINAVSKTDFADKGMGRVVRNAPAGATSGYYVYVWNRSGVSELS RVLITVETRTAGDPNNNCEFNEMPGWRVAVYGMFWQYONNERAISHSIMSK GDILRSVAYVDGAAPVFAFIEDGLSISAGADLVVMDTTFKFGATPAPECTAADVI LDRSGGFYESHSITVNDLSCKKLFATPEIYARGNOLRMIGGEGALBRNDGAT YLLTNGVDVYGGNITLPPAIDNATGELYIGKLSLNSGNALTRAKLOTTPRVSVS EFDGSKDITLTLAAHVAFARRAIDTITADADGVPVNMESGATVNTSGDSYTLVNFIT

CDS

CDS
 GVSCRTLOMKAHYRNGGLFRRSDRGYGFEDWAEVYTSKNDLPESPVGAPIMPS
 DTPVSGYALMOGQAFDKSAVPIKLAAPSGVPIPMRGWITGSPASGRVLSODDGI
 KSHTHSASSTDLGRTKTTSSFDYGTSTNNNGATHSVSGSTNSAGNHTSLANVT
 ASANGASASATRLSVVHNQNTATSSAGATHSVJSGTASAGAHHTVIGAHHTHVA
 IGSHTHTVNAAGNAENTVKNIATFNIVRLA"
 complement(3905..4066)
 /note="ORF_ID:0263#25
 similar to [SwissProt Accession Number P03774]"
 /codon_start=1
 /transl_table=11
 /protein_id="BAAL4976.1"
 /db_xref="GI:1742244"
 /translation="MAADCSDAEFAFVSDVFAAPALVAALFEDSMAALFDPASVAFVD
 AVPALEDAD"

CDS
 complement(4694..>4882)
 /note="ORF_ID:0263#26
 similar to [SwissProt Accession Number P03774]
 start codon is not identified yet"
 /codon_start=1
 /transl_table=11
 /protein_id="BAAL4977.1"
 /db_xref="GI:1742245"
 /translation="ADFTAFGVAAVSVDVLLVALLSCTIPFVVLASSATADALISSA
 RLAAVSARVADSAVLTL"
 6770..7345
 /note="ORF_ID:0264#1
 similar to [PIR Accession Number C42291]"
 /codon_start=1
 /transl_table=11
 /product="Tail fiber assembly protein G"

CDS

Alignment Scores:

Pred. No.:	63.3	Length:	16902
Score:	155.50	Matches:	121
Percent Similarity:	35.498	Conservative:	60
Best Local Similarity:	23.73%	Mismatches:	220
Query Match:	6.45%	Indels:	109
DB:	1	Gaps:	21

US-09-825-414-7 (1-486) x D90775 (1-16902)

QY	19	ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln	38
DB	3750	CGTAACGCGTCGCGGTCGACAGAACACGCGACCGCAAGAACACGCGATGATGCC	3809
QY	39	GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLys	58
DB	3810	AGCACATCAGCCCGTAGGCGGCAACCCATGCGCGCTGAT-----GCTGGGACTCA	3860
QY	59	LeuProAlaAspSerAlaAspSerAlaIleAlaValAspValHisAsnAlaGlnIle	78
DB	3861	GCACGCGCGACGACGACGCGACGCGCGCTGCTGCGCTCAGTCACGCGTCTTCC	3920
QY	79	ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGluGlyThrProAlaThr	98
DB	3921	ACGGCA-----GGAAGCGCATCAACAAAGCGCACT	3950
QY	99	IleAlaAsp---ThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrTrs	117
DB	3951	GAAGCATCAAAAAGTCTCCCTCGACAGTCTCAAAAAGCGCGCGGCGCACACTGCC	4010
QY	118	GlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnIleThrMetGlnProAla	137
DB	4011	GGTGGCGGCAAAACGTCAACAAAGAAATGCTTACGCTCCTACAAATCAGACGCACATCT	4070
QY	138	IleAsnLysGlyAspThrLeuProAlaProLeuLysProLeuThrProLeuIleSerGly	157
DB	4071	GCATCCACCGCG-----ACCAAGCAAGCATCAGAA	4100
QY	158	AlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAsp	177
DB	4101	GCTGCACCTCGCGCGGATGCGGCGCTCAAAAAGACGCGCAAAATCATCA---GAA	4157

QY	178	LeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaIle-----Ser	195
DB	4158	ACGAACGATCATCAAGCGCCAGTAGTCAGCTTCTCGGCAACGCGCGCAGGAATATTC	4217
QY	196	ValLys-----ArgHisSerProSerLeuAlaArgGln	206
DB	4218	GGAAGCGCGCAAAAACGTCCGAGACGAACGCGCAGCTTCTGAAACGCGCAGCGACAG	4277
QY	207	ValLeuAspThrGlyAlaValAlaGlnThr-----TyrSerAlaArgAsnAlaValArg	224
DB	4278	ACGCGCTCGGTCGCGGACGCTCAAAAACGCGCTCGCTGCTCCAGTCAGCGCTCA	4337
QY	225	ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly	244
DB	4338	ACAAAGCGCGGCGAGCGCTTCACCACTGCGCACCGCGCGGCAAAATCGGCAAGACGCC	4397
QY	245	ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuSerVal	264
DB	4398	GCATCGTCTGCTTCACACACCCACAAACGAGCGTGGCGAAGCCACTGAACGCGCACCGCA	4457
QY	265	GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro	284
DB	4458	GCAGCGAGGTCT-----GCTCCGACGCAAGACATCCGAAACG	4496
QY	285	LysAlaGlnLeuSerGluGlnAsnAspTrpLeuGlnAlaTyrLysAlaIleLysSerAla	304
DB	4497	AACGCAAAAGCGTCGAAACAAACG-----GCAGATCTCTCAAAAAGCGCTCGCGCATGC	4550
QY	305	SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet	324
DB	4551	TCAGCAGGTTCGCGCGCGCTCATCGCATCATCG-----GCGCTCTCTTCAAAAGATGAG	4604
QY	325	AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerLeuThrGlnAsn	344
DB	4605	GCAGCCAGCAACGCTCAGCAGCAAGAC-----AGGCGCACGCGCATCCAGCAAG	4658
QY	345	GlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlnMetAlaThrLys	364
DB	4659	GCAGCAGAGGCTGCTGCGAGT-----	4679
QY	365	AsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySer	384
DB	4680	-----GCGACGCGCGAGCTCAGAGCAAAAGTACG-----GCGGATCC	4718
QY	385	AlaAlaValPheAlaGlyTrpThrAlaAlaLeuThrThrAspProAlaValLysLys	404
DB	4719	GCGGCACGCGCGCC-----GAGACAGCACTAAACG	4751
QY	405	AlaGluSerPhe-----IleGlnAsp-----ThrValLys-----	414
DB	4752	GCAGAGATATTCATCCCGCGCGCTTGAGAGATGCAAGTACGACGAAAGGGGATTA	4811
QY	415	-----SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu	431
DB	4812	GTACACCTCAGCAGTCCGACCAACACTAGTGTGAACCGTGGCGGCAACGCCAAAGCGCA	4871
QY	432	AlaLysThrVal-----LysAspMetGlyGlyGlnAla	442
DB	4872	GTAATAATCCAGCTATGACAATGCAAGAAACGCTTCAGAAACCAAGAAAGGCGGTGAT	4931
QY	443	IleThrHisThrGlyAlaSerLeuAlaArgAsnThrValAsnAsnLeuArg-----	458
DB	4932	ATACCCGATTAAGGATGCTTCTCG---AACACATTTAAACCGGCTCAGTAAACAGACTTT	4988
QY	459	-----GlnArgProAlaArgGlnAlaAspIleGluGlnGlyGlyThrAlaAlaSerPro	476
DB	4989	GCTGATACAGCGTGTGCTGATGCGGTTAAGCTTCCTGACAGTGCACATCATCTGGA	5048
QY	477	SerGluIleProPheArgProMetArgSer	486
DB	5049	AATATATACCTGTGTGTGTATGCGTCT	5078

RESULT 12

D90774 LOCUS 18700 bp DNA linear BCR 29-MAY-1997
 DEFINITION E.coli genomic DNA, Kohara clone #263(30.5-30.9 min.).
 ACCSSION D90774 AB001340
 VERSION D90774.1 GI:1742217
 KEYWORDS Complete and shotgun sequencing; -37; IS5; racc; recE; recF; sieB; trkG; ydaB; ydaE; ydaF.
 SOURCE Escherichia coli (strain:K12) DNA, clone_11b:Kohara lambda miniset library clone:Kohara clone #263.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
 REFERENCE 1 (sites)
 AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kasai, H., Kashimoto, K., Kimura, S., Kitakawa, M., Kitagawa, M., Makino, K., Miki, T., Mizobuchi, K., Mori, H., Morita, T., Motomura, K., Nakade, S., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Sivasubraman, S., Tagami, H., Takeda, J., Takemoto, K., Takeuchi, Y., Wada, C., Yamamoto, Y. and Horiuchi, T.
 TITLE A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map
 JOURNAL DNA Res. 3 (6), 363-377 (1996)
 MEDLINE 97251357
 REFERENCE 2 (sites)
 AUTHORS Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikehota, T., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.
 TITLE The systematic sequencing of the Escherichia coli genome in Japan
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 18700)
 AUTHORS Mori, H.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
 COMMENT (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
 Colaboration Information:
 Project:
 The Japan E.coli genome DNA sequencing project
 Group:
 The Japan E.coli genome DNA sequencing group
 Members: (1995.4 - 1996.3)
 Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiuchi, T., Ikehota, T., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishori@nibb.ac.jp
 Information operator:
 Name: Hirotsada Mori
 Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
 E-mail: hmori@gtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bws3.aist-nara.ac.jp/
 Location/Qualifiers
 1..18700
 /organism="Escherichia coli"
 /strain="K12"
 /db_xref="taxon:562"

/map="30.5 min"
 /clone="Kohara clone #263"
 /clone_11b="Kohara lambda miniset library"
 /note="Nucleotide position 1416913-1435612 from the initiation site of 'thra' (0 min.).-This clone is from Kohara lambda miniset library"
 complement(1..705)
 /gene="rect"
 complement(<1..705)
 /gene="rect"
 /note="ORF_ID:0262#8
 similar to [SwissProt Accession Number P33228]"
 /codon_start=1
 /transl_table=1
 /product="Rect protein (P33)."
 /protein_id="BAA14951.1"
 /db_xref="GI:1742218"
 /translation="MTKOPPIAKADLKTGCGNRAAPAAVKSVDYISFTNPDSMEQLAAALPDMTAEKMRIRATTEIRKVPALGNCIDMSEVSAIVQCSGLGLEFSAALGNAVILPFGNKEKSGKKNVOLLIGRGMDLARSQDISLSARVREGDEFSFEGLEKLIH RQGENEDAPTVHYAVARLKDGTQGFVMTKQTELVSLSKAGNNGPWTVHMEAMK KTAIRRLKYLPSIER"
 complement(698..3298)
 /gene="recE"
 complement(698..3298)
 /gene="recE"
 /note="ORF_ID:0263#1
 similar to [SwissProt Accession Number P15032]"
 /codon_start=1
 /transl_table=1
 /product="Exodeoxyribonuclease VIII (EC 3.1.11.-) (Exo VIII)."
 /protein_id="BAA14952.1"
 /db_xref="GI:1742219"
 /translation="MSRKPLFLRKAKSSGEPDYVMAISNDFSCATIDYLIVKSG KLSISYKRAVATNPVYNDIPABEIDFTWSEBROQSKDSMTWELKGAAPDAAHSG NTNNGEDMTIEENMLPISGDELPIKMAHQSESPVTHVSRDGLQALHIAREEL PAVYALAVSKRTSLDLEIRELKLIVDTKYFPPNGSNLGLITAFEAYLVADYT DRLGILTEKMGKGNVSHITRTASGANGGNTLDGRGFVHDLTSLARDAVATGLAS MEDDLYNLHPNAKRIEITAEKPPSPVPRDKITWPGGLDYRSRAIVASVEAPIG IEVYPAVTEVLNKLVEPTHANPDEIVDIACGSSAPMPORVTEGKODDEKPPQ SCYTAVREGAETMEPDATPHHNDOPLOLDASOVNADYAKOELRAELHRAKRIIPSK NPVDDDKLILAAASRETFDGTSDPNDPRWVVGIOIRPCVYONQPTERTSPDMNPVY VQDEPELACNACGQTGDNCPCGAVNGDRTYQETDEESQVEAKENDPEMBGAEHP HENAVAGSDPHRDCSDETGEVADPVIADIEPGLIYGISNRYAHAPGISQSQDLAD TPALYLMRKNAPVDTKTTLDTLGTAFHCHVLEPEERSNRFYVAPENNRTNKGREE KAFLEMCASGTGYITAEGRKIEPLATOSWALPELGMVLVESAGHASSIYEDPETG ILCRRPDKIIPERHMTMDYKTADIORFTAYDYDYVHYODAFYSQGYEAOFGVOPF FVFLVASTTIECGRYPEITFMGEBAKLAQCEYHNRRLTSLDCLNTDEMPAIKTLSEL PRAVKEAYAND
 complement(3400..3675)
 /gene="racc"
 complement(3400..3675)
 /gene="racc"
 /note="ORF_ID:0263#2
 similar to [PIR Accession Number JS01119]"
 /codon_start=1
 /transl_table=1
 /product="Racc protein"
 /protein_id="BAA14953.1"
 /db_xref="GI:1742220"
 /translation="MTNVEAYVTDDIYHEVNLSEKRIIGYIKTENKPTPTVVDI DGSNGVKTIDEGVKKCLVHIGKNLPKAEKAEFLATLIAMKLGEL"
 complement(3750..3920)
 /gene="ydaE"
 complement(3750..3920)
 /gene="ydaE"
 /note="ORF_ID:0263#3
 similar to [SwissProt Accession Number P38394]"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA14954.1"
 /db_xref="GI:1742221"

```
/translation="MTKKIKCAVHLCKRDVESKAIERMILHFMHILSKDEPRKXKCE
ACEKDDQAHHEL"
gene
complement(3920..4141)
/gene="ydaB"
CDS
complement(3920..4141)
/gene="ydaB"
/note="ORF_ID:0263#4
similar to [SwissProt Accession Number P38393]"
/codon_start=1
/transl_table=11
/protein_id="BAAL1955.1"
/db_xref="GI:174222"
/translation="MAHHFGTDEIPROCVTPGDDVYLHFGRTYIASANNIKRKLTYIR
MLTKFTIDRMKIKVFLGRDGLPKAESW"
4583..5071
/gene="sleB"
4583..5071
/gene="sleB"
/note="ORF_ID:0263#5
similar to [SwissProt Accession Number P38392]"
/codon_start=1
/transl_table=11
/product="Superinfection exclusion protein B."
/protein_id="BAAL1956.1"
/db_xref="GI:174223"
/translation="MLDVTPLKLFANPELRMLMTYIIIFGLTLLLPKFEPTAFNA
YTEIPWFOIIVFAFSPVAISFRLRAHIOKHYSILPQRYLLRLSEKEIVAFDFL
KTCNLIITSPCNPMVKLEKRGITIQHSDSANCSYLYLTETKYSHPMKLFMNSRRRF
NR"
complement(5068..5223)
gene
/gene="ydaF"
complement(5068..5223)
/gene="ydaF"
/note="ORF_ID:0263#7
similar to [SwissProt Accession Number P38395]"
/codon_start=1
/transl_table=11
/protein_id="BAAL1957.1"
/db_xref="GI:174224"
/translation="MQKIDLGNNESLVCGVFPNODGTFTAMTYTTSKFTKTTGARWR
LEKHTVS"
7895..8641
/note="ORF_ID:0263#10
similar to [SwissProt Accession Number P07905]"
/codon_start=1
/transl_table=11
/product="DNA replication protein DnaC."
/protein_id="BAAL1958.1"
/db_xref="GI:174225"
/translation="MKNIATGDLVERIRRLAPSHVAPKTYVAEMREMOISGOKRCE
EINRONROLRYEKILINRSIOPIOLHRKCSFSNOYONEGORVLAISOAKSTADELMTCGT
NPAFSGKRGTCNNHAAAIAGNRLKLDGOTVYIVYADVMSALHAISTDDOSEKFLRE
LCEVDLVLDIEIGIRETKNEQVYLHQTVDRTASMSRVGLMTNLNTEAMKTLIGERI
MDRMTNGCRGWNFMWESMRPNVOPGIAK"
9251..9556
CDS
Alignment Scores:
Pred. No.: 72.1 Length: 18700
Score: 155.50 Matches: 121
Percent Similarity: 35.49% Conserved: 60
Best Local Similarity: 23.73% Mismatches: 220
Query Match: 6.45% Indels: 109
DB: 1 Gaps: 21
US-09-825-414-7 (1-486) x D90774 (1-18700)
QY 19 ArgTThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln 38
||| ||||| :::::::::::::::::::: |||||
Db 15303 CGTAACGCGTCCGCGGTGGACAGAACACGCCGCGGAGAGACGACGATGATGCC 15362
QY 39 GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLys 58
||| :::::::::: |||||
Db 15363 AGCAATCAGCCCGTGGAGCGGACCAACCATGCGGCTGAT-----GCTGCGGACTCA 15413
```

```
QY 59 LeuProAlaAspSerAlaAspGluGlnAlaAlaValAspValHisAsnAlaGlnIle 78
||| :::::::::: |||||
Db 15414 GCACGGCGGACCCAGCAGCTCAGCCGGACAGCCGCGCTCGTCCAGTCAAGCTTCC 15473
QY 79 ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGluThrProAlaThr 98
||| :::::::::: GAAACGGCATCAACAAAGCCGACAT 15503
Db 15474 AGCGCA-----
QY 99 IleAlaAsp---ThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSer 117
||| :::::::::: |||||
Db 15504 GAAGCATCAAAAGTCTCGCTCGGATCGCAAGTCTCAAAAGCCGGCGGACCACTGCGC 15563
QY 118 GlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnThrMetGlnProAla 137
||| :::::::::: |||||
Db 15564 GTGCGCGGCAAAACGTCAGAAACGAATGCTTCAAGCTCACTACAAACAGCAGCCATCT 15623
QY 138 IleAsnLysGluAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGly 157
||| :::::::::: ACACGAAGGCATCAGAA 15653
Db 15624 GCATCCACCGCG-----
QY 158 AlaLeuSerGluAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAsp 177
||| :::::::::: |||||
Db 15654 GCTGCGACCTCGGCGCGGATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 15710
QY 178 LeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAla-----Ser 195
||| :::::::::: |||||
Db 15711 ACAGACGATCATCAAGCCGCACTAGTCAAGCTTCTCGGCAACGCGGCGGCGGCAATTCC 15770
QY 196 ValLys-----ArgHisSerProSerLeuAlaArgGln 206
||| :::::::::: |||||
Db 15771 GCAGAGCGGCAAAACGTCGAGAGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 15830
QY 207 ValLeuAspThrGlyValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArg 224
||| :::::::::: |||||
Db 15831 AGCGCTCGGCTCGGCGGCGGCTCAAAACAGCGGCTGCGCTCGTCCAGTCAAGCGGCTCA 15890
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGluValAlaLysLeuGly 244
||| :::::::::: |||||
Db 15891 ACAGAGTCCGGCGGCGGCTCAGCAGTCCAGCAGCGGCGGCGGCGGCGGCGGCGGCGG 15950
QY 245 ValSerMetAlaGlyLysLeuAlaAlaAsnAlaGlyPheGluLysAsnArgLeuSerVal 264
||| :::::::::: |||||
Db 15951 GCATCGTCTGCTTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 16010
QY 265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGluLysAspLysGluPro 284
||| :::::::::: |||||
Db 16011 GCAGCGAGGTCT-----GCTTCGCGACGGAAGCATCCGAAACG 16049
QY 285 LysAlaGlnLeuSerGluGluAsnAspTyrLeuGluAlaThrLysAlaIleLysSerAla 304
||| :::::::::: |||||
Db 16050 AACGCGAAGCGTGGGAACACAGC-----GCAGAAATCTCAAAACGCGTCCGATCG 16103
QY 305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet 324
||| :::::::::: |||||
Db 16104 TCGACAGTTCGGCGGCGGCTCATCGCATCGCATCG-----GCTCTGCTTCAAAAGTAG 16157
QY 325 AlaThrAspAlaMetCylAlaValArgSerLeuValSerAlaSerLeuThrGlnAsn 344
||| :::::::::: |||||
Db 16158 GCGACGACGACAGACGTCAGCAGGAGAGAGC-----AGCGCCACGACGCGATCCAGAG 16211
QY 345 GlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLys 364
||| ::::::::::
Db 16212 GCGACAGAGGCTGCTGCGAGT----- 16232
QY 365 AsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySer 384
||| :::::::::: |||||
Db 16233 -----GCGACGCGGCGGCGGCTCAGACGAAAGTAGTC-----GCGGAATCC 16271
QY 385 AlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLys 404
||| :::::::::: |||||
Db 16272 GCGGCAACGCGCGC-----GAGACAGCAGCTCAAAACG 16304
```

Oy	405	Alagiusserpine-----lleginasrpe-----ThValLys-----	414
Db	16305	GCGAGGATATTGCATCCGCCGTGGCGTTGAAGATGCAAGTACACAGCAAAGGGGATA	16364
Oy	415	-----SerThAlaserSerThrThrcgLyTyrrValAlAspGlNthrVallysLeu	431
Db	16365	GTACAGCTCAGACATGCGCACCAACAGTAGTCGTGAAAGCGTGCGGCAACGCGCAAGGCA	16424
Oy	432	AlalySThrVal-----LysAspMetcLyglyLuAla	442
Db	16425	GTAAATATGACCTTAATGACAMTGCAGAAACGCTCTGCAGAAAAGACCAGAACGCGCTGAT	16484
Oy	443	IleThrHISThrGlyAlASerLeuIrrgaSnThrValAsnAsnLeuArg-----	458
Db	16485	ATACCCGATMAGGGATGCTTCTCTG---AACAACATTAACCGCGGTACGTAAAAACAGACTTT	16541
Oy	459	-----GlnArgProAlaArgGluAlaAspIlegIugIugLyglyThraAlaIaserPro	476
Db	16542	GCTGATAGCGGCGTATGCGTATGCTGGCGGTAAACGCTCCGCGAGGTGCAACATCTGGA	16601
Oy	477	SergIulleProPharPromeChrgerSer	486
Db	16602	AAATATPACCGTGTGTGTTATGCGTCT	16631
RESULT 13			
LOCUS	AF317649	5329 bp DNA linear BCT 10-NOV-2000	
DEFINITION	Rhodobacter sphaeroides flagellar hook-associated protein 3 (flgK) and flagellar hook-associated protein 3 (flgL) genes, complete cds.		
ACCESSION	AF317649		
VERSION	AF317649.1	GI:11136618	
KEYWORDS	.		
SOURCE	Rhodobacter sphaeroides.		
ORGANISM	Rhodobacter sphaeroides.		
REFERENCE	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;		
AUTHORS	Rhodobacter.		
TITLE	1 (bases 1 to 5329)		
JOURNAL	Gonzalez-Pedrajo,B., De La Mora,J., Ballado,T., Camarena,L. and Dreyfus,G.		
REFERENCE	Isolation and Complementation of a Flagellar P-ring Mutant of Rhodobacter sphaeroides		
AUTHORS	Unpublished		
TITLE	2 (bases 1 to 5329)		
JOURNAL	Gonzalez-Pedrajo,B., De La Mora,J., Ballado,T., Camarena,L. and Dreyfus,G.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (31-OCT-2000) Genetica Molecular, Instituto de Fisiologia Celular, UNAM, Circuito Exterior, Ciudad Universitaria, Mexico Distrito		
FEATURES	location/Qualifiers		
source	1..5329		
gene	/organism="Rhodobacter sphaeroides"		
CDS	/db_xref="taxon:1063"		
	1..4092		
	/gene="flgK"		
	1..4092		
	/gene="flgK"		
	/note="HAP 1"		
	/codon_start=1		
	/transl_table=1		
	/product="flagellar hook-associated protein 1"		
	/protein_id="AAG31286.1"		
	/db_xref="GI:11136619"		
	/translation="MSIIIDIVARSGLLSYRSALSVTAENIANVNTGGYVRREAVLAQVP GGQISPTSAATSGGVVEDYRAFRADFLVAORITAFAGAASAATLTDTTGAAGLESLFL AGASVPEALTGTGFDALNGCATPRDIALROYVLAACOTIASSTTYASGLACIRDEV TGLAGQTVASRSTLRDLAEINGRMTAQOQSALNPILDERDRLLIGLSRVGVASFDF ALGRTVTIGSPAGGPILLLEGTAATSLGVNAAGPLVLEYTRAGATTLSROITGSSLGG YASLSIAVDNMAAEILDALARVLAKMTVAEMTVAAOGIDGARGGDLEFLEGOVADACN RGTCGASVAVEDETAFAGPLTLHDAAAGLHWAHADAGGTILGSCAETLATLGHLWMT GAAMDGRITVSPROGARDLMRELTPTORIAAAPITYVSANSNASQAQSPVPTS TIPTATGSLGELTDAGCADANATILRAGVVGHTIPAGTOSVTLASGSOQDEFLISDPA AAGVTTLTSFIGTIVHSFDLSALGATDPATLAAALNLNAGAAVSTGGOTLAEIGVASGST		

Db	2235	GCACCAACT---	CTCCGCCGCTGGCGATCAGCGCGCGCTGC	CGGGGGGGCGCATGATG	231111
Qy	198	GHisserProSerLeuAlaIarV-	GLnVal-----	LeuAspT	210
Db	2195	CCCGACCGCGGTACGATCGAGCGCGCCCGACGCGACGCTCA	CGGTCCGCGTCA	CGCGCG	2254
Qy	134	rProAspArgLeuHisAspAlaMetala-----	AlaSerValys-----	Ar	198
Db	2155	CCCCACCGCGGTACGATCGAGCGCGCCCGACGCGACGCTCA	CGGTCCGCGTCA	CGCGCG	2254
Qy	14	pGlnValGlyThrLysMetCetaSparGAlaIarhGlyAspRLeuHisTyrLeuSerAlaSe	184		
Db	2154	GCAGGCCAT-CAACGCGTTCGGGGCGCGGATCGAGGC-----	CGAGG	2194	
Qy	14	uProAlaProLeuLysProLeuThrProLeuHisSerGlyAlaLeuSerGlyAlaMetas	164		
Db	216	TCCCTCCCGGTGAG-----	CTGCCGTGGGTGCTCGCGCGCGGACATGCG	2155	
Qy	14	pGlnValGlyThrLysMetCetaSparGAlaIarhGlyAspRLeuHisTyrLeuSerAlaSe	184		
Db	2052	-----TCTGCTCTCTGCACTCCCGGCTCCCGACCTCGACGCTCGAAGCGGGGGGGCT	2105		
Qy	85	ArgAlaSerArgLeuHisPheGlnGlyGluThrProAlaThrIleAlaAspThrPheAla	104		
Db	1956	CGGCTTCGCGGACCGGAGTCCGGGGCGCGATCTGCGGGGACGCTGCTCGCT	2015		
Qy	105	LysAlaGlnLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrPro	124		
Db	2016	TCACTCCCGCCCA-----	TCCGTGGCGCGGACGTTTCGGGCT	2051	
Qy	125	PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGly-AspTyrLe	144		
Db	2052	-----TCTGCTCTCTGCACTCCCGGCTCCCGACCTCGACGCTCGAAGCGGGGGGGCT	2105		
Qy	14	uProAlaProLeuLysProLeuThrProLeuHisSerGlyAlaLeuSerGlyAlaMetas	164		
Db	216	TCCCTCCCGGTGAG-----	CTGCCGTGGGTGCTCGCGCGCGGACATGCG	2155	
Qy	14	pGlnValGlyThrLysMetCetaSparGAlaIarhGlyAspRLeuHisTyrLeuSerAlaSe	184		
Db	2154	GCAGGCCAT-CAACGCGTTCGGGGCGCGGATCGAGGC-----	CGAGG	2194	
Qy	134	rProAspArgLeuHisAspAlaMetala-----	AlaSerValys-----	Ar	198
Db	2155	CCCCACCGCGGTACGATCGAGCGCGCCCGACGCGACGCTCA	CGGTCCGCGTCA	CGCGCG	2254
Qy	198	GHisserProSerLeuAlaIarV-	GLnVal-----	LeuAspT	210
Db	2235	GCACCAACT---	CTCCGCCGCTGGCGATCAGCGCGCGCTGC	CGGGGGGGCGCATGATG	231111

```

Oy 210 hr---GlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaP 229
Db 2312 CGCTGGCGCTGGCGCTCAATGCGGTGACCGCGCCACCGCGGTGGCGCGGAGCTGTGCG 2371
Oy 229 roAlaLeuAlaSerArgProAlaValGln---GlyAlaValAlaSerLeuGlyValSerMeta 248
Db 2372 CGGAGCGGGCGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2427
Oy 248 laGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgA 268
Db 2428 ----GGCTCTCGCCACACACGGCGGGCGCGCGCTGACCTCTGACGGCACGGACGGAGG 2482
Oy 268 spHisGlnArgGlyGlyAlaValLeuValLeu---GlyLeuLysAspLysGluProLysAlaG 287
Db 2483 GCAGTCCCGGACGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2533
Oy 287 InLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaLysSerAlaSerTyrS 307
Db 2534 GCTTCACGGCGGAG-----CTGCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2566
Oy 307 enGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrA 327
Db 2567 GCGGTCTTCGCGCGCATGTGGGC-----GGCGTGGCGGAGATGCGCGCGCGCTGG 2614
Oy 337 spAlaMet-----GlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn- 344
Db 2615 ATCCGATAGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2674
Oy 345 -----GlyLeuAlaLeuAlaGlyLysPheA 353
Db 2675 ATGCGTTCACCTACGATCCCGCTCTGACGGGGCGGGCTCTGCGCGACGACCTCTGCG 2734
Oy 353 laGlyValGlyLysLeuGlnGluMetAlaThr-----LysAsnLeuThrAspProA 370
Db 2735 CGCAGCGCGCTCCCGCGAGATGCGGATGCGGAGCGGCGGACCGGAGCTGACATCGATG 2794
Oy 370 laThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySer----- 384
Db 2795 CCGCGCGCGCGCGCTGAGCGACGGCGCGCGCTGCGCTCTCTCTCTCTCTCTCTCTCT 2854
Oy 385 -----AlaAlaValPheAlaGlyTrpThrAlaAlaLeuThrAsp 400
Db 2855 GCGCGGAGCGCGCGCGCGCGCTGACCGCGCGCGCGCGCGCGCTGCGCGCGCTG-----C 2905
Oy 400 roAlaValLysLysAlaGlnSerPheIleGlnAspThrValLysSerThrAlaSerSert 420
Db 2906 CCGCGGATGGCGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2965
Oy 420 hrThrGlyTyrValAla-----AspG 427
Db 2966 CGCGTGGGGCGGTGCGGTGCGGTGCGGCGGAGCGGAGCTGCTACGGCGCGCTTCGACG 3025
Oy 427 InThr-----ValLysLeuAlaLysThrValLysAspMetGly----- 439
Db 3026 CGACGACCGTCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3085
Oy 440 ----GlyGlyAlaLeuThrAsnThrGlyLysSerLeuArgAsnThrValAsnAsnLeuA 458
Db 3086 AGAGCGGGCGCGCGCGCTTCGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 3145
Oy 458 rgGlnArgProAlaArgLysAla-----AspIleGluGluGlyGlyT 472
Db 3146 CCGGTACGCGCGCGGATCCCGCGCGCTGCGCGCGAGCTGACATCGAGCTGCGGCGGA 3205
Oy 472 hrAlaAlaSer 475
Db 3206 CCTCTACAGC 3216

RESULT 14
LOCUS AL646078 203050 bp DNA linear BCT 07-DEC-2001
DEFINITION Ralstonia solanacearum GM11000 megaplasmid, complete sequence;
segment 3/11.

```

```

ACCESSION AL646078 AL646053
VERSION AL646078.1 GI:17430778
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 203050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Couzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavi,M., Moisan,A., Robert,C., Saurin,M., Schlex,T.,
Siguer,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 203050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URCV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.

COMMENT location/Qualifiers
FEATURES
source 1..203050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
/complement(147..1499)
/gene="RSP0310"
/note="RS05460"
/complement(147..1499)
/gene="RSP0310"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon-usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not-experimental
/transl_table=11
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CAD17461.1"
/db_xref="GI:17430779"
/translation="MNAPRKNTDRAEALILRNASTHEPANPCWIPDAAAGLHPGR
WRAPFAIFGGQALSLGSAITQVLLMTIDTGSVALATAGLALIPDAVISPGR
CTFADRYSRRLTMVADGISLMLLILFELRILMHAHYVMALISAQAQARA
AAASVTMLVREGFTVRAAGLQNSQSLTLVYAALGALATVMPYIGALGIDVATALL
GIAPLCCRIPOACVSSGQKTLMPARREGVDLWKPGLVLLGAVVLTIPTF
TLVPLVKTFRGGAQVAVLMEGLSGVAGLVVAMAPRVQVAVILGFAVSCVA
LALTALPGLSFGVAVMVMVISTGFYFGNPLTALLQTIYPNHLQGRVSLNTTNG
LAAPVGLALFTPLCEVIGVRLFAIAGLGAASLAGFLPSLRLDRRPSRSARD"
/complement(1531..2961)
/gene="optM"
/note="RSP0311: RS05459"
/complement(1531..2961)
/gene="optM"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon-usage
predicted by Homology

```


OY	70	AlaValaSerValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSerArgLeu	89
Db	94863	GCC-----	94874
OY	90	HisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeu	109
Db	94875	-----	94892
OY	110	AspArgLeuAlaIleThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer	129
Db	94893	GACGGGCCCAAGGCCAGCGCGCGCCGCAAC-----	94922
OY	130	LeuLeuGlnIleTyrMetGlnProAlaIleAsnLysGlyAspTyrPleuProAlaProLeuLys	149
Db	94923	---GATCAGACAGCGCGCTCAGAGCCGACCGCAATGCCGACGGCGCGCCGCGC-----	94970
OY	150	ProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys	169
Db	94971	-----ACGCGCGCCCTTCGCGCCACGAGATGACAGACAGAGCTTCACAG	95012
OY	170	MetMetAspArgAlaThrClyAspLeuHisTyrIleuSerAlaSerProAspArgLeuHis	189
Db	95013	-----	95036
OY	190	AspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAsp	209
Db	95037	GACAAAGACCGCGCGCGCGCGCGCGCGCTGCATCCCGCG-----	95075
OY	210	ThrGlyValAlaValGlnThrTyrSerIleArgAsnAlaValArgThrValLeuAlaPro	229
Db	95076	GCACAGCTGGCGCGCGCAATGAGACCCGCGCGCCATGTCACACAGCGCGCGCTCCG	95135
OY	230	AlaLeuAlaSerArgProAlaValGlnLysAlaVal-----AspLeuGlyValSerMet	247
Db	95136	-----GTCCTCCAGCGCGCGACGAGGTCCGAGCGGCGAGACGATGCCGCGCGCAGCTC	95189
OY	248	AlaGlyGlyLeuAlaAlaAsnAlaGlyIleGlyIleAsnArgLeuLeuSerValGlnSerArg	267
Db	95190	GCGGCGGCGCAAGCGCGCGCGCGCCACCAACGACGCGCGCGCTGCGCGCCCTGAGCCG	95249
OY	268	AspHisGlnArgGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGln	287
Db	95250	GCCAAATGCCGCGCGCGCAATTGACC-----GATCCGCGCGCCACC	95291
OY	288	LeuSerGluGluAsnAspTyrPleuGlnAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db	95292	GTCGCGCCAGCGC-----CAGCGTGGCAGACGCGAGCGCTCCGATGCCAGCACC	95330
OY	308	GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu--AspMetAlaThr	326
Db	95331	GCGGCGCGCGCAAGAGAGCCACGCGCGCAAGCCGACGCGCTGCGCGATGCCAGCACC	95390
OY	327	AspAlaMetGly-----	333
Db	95391	GTCGCGCCAGCGCTGCGCGCGCGACGCGCGGCAACACCGCAATATGCCCGCGCGCTGCGC	95450
OY	334	SerLeuValSerAlaSerSerLeuThrGln-----AsnGlyLeuAlaLeu	348
Db	95451	CACCTGGCGCGCATCTCTCTCGCGCGAGGCGCGCGCGCGCGACGACGCTGAGCGCGCGC	95510
OY	349	AlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAsp	368
Db	95511	CACGCGACCGAAGCGCGCGAGCGACGACGCGCAAGCGCGACCGCGCGCCGCGCCAGCGCCAG	95570
OY	369	ProAlaThrLysAlaAlaValSerGlnLeuThrAsn-----LeuAlaGlySer	384
Db	95571	ACCGCGCCCAACGCGGACACCGACGCGCAACCCCAACGACGCGCGATGCCAGGCGCGCC	95630
OY	385	AlaAlaValPheAlaGlyTyrThrThrAlaIleLeuThrThrAspProAlaValLysLys	404
Db	95631	GCGCGCAAGTTCGATGGGTCTCTGACGCGCGCC-----AACGGGATGCTGCTCCACCC	95684

Oy	405	AlaGluSerPheIleGlnAspThrVallysserThrAlaSerSerThrThcglyTyrral	424
Db	95685	GCGCCACCTTC-----ACGTGGCGCGCGCGGTACCAGCCGCACAGCAGC	95735
Oy	425	AlaAspGIntnTrVallyLeuAlaIysThrVallysaSmetGlyGIuAlaIlethr	444
Db	95736	GCCGGCGTACACCCGCCAGCGCATCCTGCCAGACTTCGGGATGCCGCCGCGCGC	95795
Oy	445	HistHrGlyAlaSerLeu-----ArgasnThrValasnaSnu	457
Db	95796	CACACCAACGCGAGCACCTGGCTACATGAGGTCATCGCCAGTCTCCGCCGAACTG	95855
Oy	458	ArgGlnArgProAlaArg-----GluAlaAspIleGlnGlyly	471
Db	95856	CAGTTGACCCCGCGCGAGCTCGCGGCTGCATGTCAAGTCGAGATGACACAGCGCCT	95915
Oy	472	ThrAlaAlaSerProSerGluIleProPheArgPrometArger	486
Db	95916	GTCATGCGCAGCT--TCCTCTGTCTACGACACAGCGCCGTGGCGGAAC	95958
RESULT 15	MTCY493	40790 bp DNA linear BCF 03-AUG-2001	
LOCUS	Mycobacterium tuberculosis H37Rv	complete genome; segment 63/162.	
DEFINITION	295844 AL123456		
ACCESSION	295844.1 GI:3250713		
VERSION			
KEYWORDS			
SOURCE			
ORGANISM	Mycobacterium tuberculosis H37Rv. Mycobacterium tuberculosis H37Rv Bacteria; Firmicutes; Actinobacteria; Actinobacteriales; Actinomycetales; Corynebacterineae; Mycobacteriaceae;		
REFERENCE	Mycobacterium tuberculosis complex. 1 (bases 1 to 40790)		
AUTHORS	Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C., Hayes,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E., Tetralis,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T., Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S., Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J., Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A., Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S., Squires,R., Squires,R., Sulston,D.E., Taylor,K., Whitehead,S. and Barrell,B.G.		
TITLE	Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence		
JOURNAL MEDLINE	Nature 393 (6685), 537-544 (1998)		
PUBMED	98295987		
REFERENCE	2 (bases 1 to 40790)		
AUTHORS	Parkhill,J.		
TITLE	Direct Submission		
JOURNAL	Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75712 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk		
COMMENT	On Jun 24, 1998 this sequence version replaced gi:2131042.		
	Notes:		
	Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers. Gene prediction was based on a Hidden Markov Model of TP genes implemented in TParse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C. CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.		

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:18:04 ; Search time 238.754 seconds

(without alignments)
4584.089 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412

Sequence: 1 MHNRVQOPVATATDSFRT.....IEEGTASPEIFRPMRS 486

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL=frame+2n.model -DEV=xlp
-O=/cgn2.1/USPTO/US09825414/runat_28012003_161741_26075/app_query.fasta.1.1294
-DB=N.Geneseq_101002 -QFMT=fastap -SUFIT=p2n.rng -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414 -ECGN.1.1.103 &runat_28012003_161741_26075 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUENT -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMCOUT=120
-WARN_TIMCOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7
```

Database :

```
1: N.Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2412	100.0	1461	22	AAD20408	P. syringae pv. tom
2	2412	100.0	1461	22	AAFS5683	Nucleotide sequenc
3	2412	100.0	30365	22	AAD20405	P. syringae pv. to
4	1869.5	77.5	1464	22	AAD20438	Pseudomonas syring
5	168	7.0	4403765	22	AA199683	Mycobacterium tube
6	164	6.8	4411529	22	AA199682	Mycobacterium tube
7	148.5	6.2	4645	24	AA196881	Lactobacillus rham
8	144.5	6.0	6744	13	AAO29471	Extracellular fact
9	143.5	5.9	11883	21	AAAI0263	Bordetella pertuss
10	143	5.9	1801	20	AAZ20203	Mycobacterium tube
11	142.5	5.9	7545	22	AAK51943	Human polynucleoti
12	142.5	5.9	7615	22	AA158449	Human polynucleoti
13	142.5	5.9	7657	22	AA158468	Human polynucleoti
14	141	5.8	2466	23	ABL27385	Drosophila melanog
15	141	5.8	53789	19	AAV21187	Amecolopsis medi
16	139	5.8	1821	23	AA581976	DNA encoding novel
17	138.5	5.7	4411529	22	AA199682	Mycobacterium tube
18	137.5	5.7	2700	22	AAD02810	HFICU08 CDNA clone
19	137.5	5.7	12036	11	AAO04668	FHA structural gen
20	137.5	5.7	32768	19	AAV52204	Streptococcus pneu
21	135.5	5.6	2550	19	AAV52497	Streptococcus pneu
22	135	5.6	2172	21	AAZ54239	Neisseria meningit
23	135	5.6	29069	21	AA81497	N. meningitidis pa
24	135	5.6	349980	21	AAAF21609	Neisseria meningit
25	134.5	5.6	4632	23	ABL25113	Drosophila melanog
26	134.5	5.6	9365	23	ABL25112	Mycobacterium spec
27	133.5	5.5	5036	22	AAD28342	Mycobacterium tube
28	133.5	5.5	5036	22	AAD13450	Human novel cytol
29	132.5	5.5	1491	22	AA559820	Lymphoma associate
30	132.5	5.5	1924	24	ABK72305	DNA encoding lymph
31	132.5	5.5	1924	24	ABK72323	Streptomyces noirs
32	132.5	5.5	65140	22	AAD17184	Streptomyces noirs
33	132.5	5.5	125401	22	AAD17186	Streptomyces noirs
34	132.5	5.5	4403765	22	AA199683	Mycobacterium tube
35	132	5.5	47981	22	AAE30757	Mycobacterium tube
36	131.5	5.5	6258	23	AA552002	Mycobacterium tube
37	131.5	5.5	17388	23	AA554875	Staphylococcus aur
38	131.5	5.5	77536	21	AAAI4651	Nucleotide sequenc
39	131	5.4	1800	24	ABK14137	DNA encoding anti
40	131	5.4	2816	23	ABL12469	Drosophila melanog
41	130.5	5.4	44377	18	AA78508	Platenolide synth
42	130.5	5.4	44377	18	AA78508	Platenolide synth
43	129.5	5.4	3606	23	ABL09625	Drosophila melanog
44	129.5	5.4	7035	23	AA554978	Staphylococcus aur
45	129.5	5.4	9345	23	ABL09624	Drosophila melanog

ALIGNMENTS

RESULT 1

AAD20408 standard: DNA; 1461 BP.

AAD20408;

03-JAN-2002 (first entry)

P. syringae pv. tomato (Pto) DC3000 CEL open reading frame 5 (ORF5) DNA.

Conserved Effector Loc1; CEL; cytostatic; antibacterial; gene therapy;

KW Exchangeable Effector Loc1; EEL; disease resistance; transgenic plant;

KW eukaryotic cell death; cancer; ds.

OS Pseudomonas syringae.

XX key Location/Qualifiers

PH CDS 1..1461

FT /*tag= a
FT /product= "Pto DC3000 CEL ORF5 protein"
XX
XX
XX WO200175066-A2.
XX
XX 11-OCT-2001.
XX
XX 03-APR-2001; 2001WO-US10698.
XX
XX 03-APR-2000; 2000US-194160P.
XX 11-AUG-2000; 2000US-224604P.
XX 17-NOV-2000; 2000US-249548P.
XX
XX (CORR) CORNELL RES FOUND INC.
XX (JUNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
XX (UYNE-) UNIV NEBRASKA.
XX
XX COLLIER A, Alfano JR, Charkowski AO;
XX
XX WPI; 2001-639361/73.
XX P-PSDB; AAE12573.
XX
XX
XX New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants
XX
XX
XX Claim 1; Page 21; 217pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 DNA.
XX
XX Sequence 1461 BP: 311 A; 440 C; 451 G; 259 T; 0 other:
SQ

Alignment Scores:
Pred. No.: 3.8e-169 Length: 1461
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-825-414-7 (1-486) x AAD20408 (1-1461)

QY 1 MethisileasnaArgArgValGlnGlnProValThrAlaThrAspSerPheArgThr 20
DB 1 ATGCACATCAACCGAGCGGTCCACACCGCTGTGACGCGACGATAGCTTTGGGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnGlnArg 40
DB 61 GCGTCGACGCGCTCTTGCTCCACGCTCTGTGCGATCTGTGAGCTCCGATCGACACACG 120

QY 41 GluIleasnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
DB 121 GAGATAAATGCGATTGCCGATTAAGTACCTGACAGATCATGTGTGGCTGCACATAACTGCCG 180
QY 61 ProAlaAspSerAlaAspGlyGlnAlaValAspValHisAsnAlaGlnIleThrAla 80
DB 181 CCGGCCGATTGCGCTGATGGCGCAAGCTGACGTGACACAAATGCGCAATCATCTCCG 240
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
DB 241 CTGATCGAGACCGCGCCACCGCTGACCTTCGAAAGGGAAACCCCGGCAACATCGCC 300
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu 120
DB 301 GACACCTTCGCCCAAGCGGAAAGCTGACCGATTGGCGAGACATCATCAGGCGCGTTG 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140
DB 361 CGGGCGACGCGCTTGGCCATGGCTCGCTTGCCTCAGTACATGACACCTGCGATCAACAG 420
QY 141 GlyAspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
DB 421 GCGCATTTGGCTGCCGCGCTCCGCTCAACCGCTGACCCCGCTCATTTCCGAGCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr 180
DB 481 GCGCCCATGAGACCAAGTGGCGCACCAAGATGATGACCCCGCGACGCGGATCTGCATTAC 540
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
DB 541 CTGACGCGCTCGCGGACAGGCTCCACAGATGCGATGGCCGCTTCGCTTAAGGCCACTCG 600
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
DB 601 CCAAGCCTTGGCTCGACAGATGCTTGACACGCGGGGTGGCGTTACAGACTACGCGCGCG 660
QY 221 AsnAlaValAlaThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
DB 661 AACGCGGACGACGACGATGTTGGCTCCGACAGCGGCGTCCAGCCCGCTGCACAGGTCCT 720
QY 241 ValAspLeuGlyValSerMetAlaGlyLysLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
DB 721 GTGGACCTTGGTGTATTCATGATGGCGGGTGTCTGCTGCCAACGCGCTTTGGCAACCG 780
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
DB 781 CTGCTCAGATGTCAGTACGCGGTATCACACGCGTGGCGGTGATTTAGCTCGGTTTAAAG 840
QY 281 AspLysGluProLysAlaGlnLeuSerGluGlnAsnAspTyrPleuGluAlaTyrLysAla 300
DB 841 GATTAAGAGGCCCAAGCTCAACTGACGAAAGAAAAGCATGCTCGAGGCTTATTAACCA 900
QY 301 IleLysSerAlaSerTyrSerGlyValAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB 901 ATCAATTCGCGCAGCTACTCGGGTGGGCTCAACGCTGGCAACGCGATGGCCGCTG 960
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
DB 961 CCACCTGATATGCGACCGACCGCATGGGTGGTGAAGAAGCGTGTGACGCGTCCAGC 1020
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyLysPheAlaGlyValGlyLysLeuGlnLys 360
DB 1021 CTGACCCCAAAAGCGTCTGGCGCTGGCGGGTGTTCGACAGGGGTGAGCAAGTTGGACGAG 1080
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGluLeuThrAsn 380
DB 1081 ATGGCGAGAAATAATACCGACCGCGGACCAAGCGCGGGGTAGTCAGTTGACCAAC 1140
QY 381 LeuAlaGlySerAlaAlaValAlaPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro 400
DB 1141 CTGGCAGAGTTGCGACGCGCTTTTCGACAGCTGACACGCGCGCTTACAAACGATCC 1200
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420

```
|||||
Db 1201 GCGGTGAAAAAGCCGAGTGTTCATACAGACAGCGTGAATCGATCGATCACAGACC 1260
QY 421 ThGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 1261 ACAGGCTACTACCTACCGACGACGCTCAAACTGCGCAAGACCGTCAAAAGCATGGCGGG 1320
QY 441 GLuAlaIleThrHisThrGlyAlaSerIleuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 1321 GAGGCGATCACCCATACCGGCGGCGAGCTTGGCAATACGGTCAATACCTGCGTCAACGC 1380
QY 461 ProAlaArgGluAlaAspIleGluGlyGlyThrAlaLaseProSerGluIlePro 480
Db 1381 CCGGCTCGTGAAGCTGATATAGAAAGAGGGGCGACGCGGCTCTCCAACTGAAATACCG 1440
QY 481 PheArgProMetArgSer 486
Db 1441 TTTCGGCTATGCGGTGC 1458

RESULT 2
ID AAF55683 standard; DNM: 1461 BP.
AC AAF55683;
XX
DT 11-JUN-2001 (first entry)
XX
DE Nucleotide sequence of a HopTox protein of Pseudomonas syringae.
XX
KM HIV; tat protein; effector protein; transduction domain; HopTox protein;
XX
OS Pseudomonas syringae.
XX
EH Key Location/Qualifiers
FT CDS 1..1461
FT /tag= a
FT /product= "HopTox"
XX
PN WO200119393-A1.
XX
PD 22-MAR-2001.
XX
PF 13-SEP-2000; 2000WO-US24977.
XX
PR 13-SEP-1999; 99US-0153507.
XX
PA (CORR ) CORNELL RES FOUND INC.
XX
PI Collier A, Beer SV;
XX
DR MPI: 2001-257850/26.
DR P-PSDB: AAB67677.
XX
PT Delivering effector proteins into target cell for use in protein
PT therapy, involves introducing effector protein fused to protein
PT transduction domain of human immunodeficiency virus Tat protein, into
PT target cells -
XX
PS Example 2; Page 22-23; 43pp; English.
XX
CC The present sequence encodes a HopTox protein of Pseudomonas syringae
CC pv. tomato DC300 CEL.. This is an effector protein, which is used in
CC the method of the invention. The specification describes a method for
CC delivering effector proteins into a target cell. The method comprises
CC introducing an effector protein fused to a protein transduction domain
CC of a human immunodeficiency virus (HIV) tat protein into the target
CC cell. The method is used for delivering a heterologous effector protein
CC such as an effector protein produced by bacterial plant pathogen, animal
CC pathogen or a rhizosphere bacteria, or a protein secreted and/or
CC delivered into eukaryotic cells by a type III secretion system or a
CC hypersensitive response elicitor, an avirulence protein, a
CC hypersensitive response and pathogenicity-dependent outer protein, a
```

```
CC virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC The method is useful for delivering effector proteins for use in
CC pharmaceutical, insecticidal, fungicide, herbicide and other
CC applications.
XX
SQ Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other;
Alignment Scores:
Pred. No.: 3 8e-169 Length: 1461
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-825-414-7 (1-486) x AAF55683 (1-1461)
QY 1 MethiSlleAsnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThr 20
Db 1 ATGCACATCAACCGACCGCGTCCAAACACCGCTGTACTCGACGATACCTTCGGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnArg 40
Db 61 GCGTCCGACGCGTCTCTTGGCTCCAGCTGTGCGATCTGCAGTCCGATCGAACAGC 120
QY 41 GluIleAsnAlaIleAlaAspTyrIleuThrAspHisValPheAlaAlaHisLysLeuPro 60
Db 121 GAGATAAATGCGATGCGCGATTCCTGACAGATCATGTCTTCCGCGCTAAACGTCCG 180
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAlaAspValHisAsnAlaGlnIleThrAla 80
Db 181 CCGGCGATTCGGCTGATGGCCAAAGCTGCAGCTGACCTACACAAATGCGGATCAGTCCG 240
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
Db 241 CTGATCGAGAGCGCGCGACCGCTGCATCTGAAAGGGGAAACCCGCGCAACATCGCC 300
QY 101 AspThrPheAlaLysAlaGlyLysLeuAspArgLeuAlaThrThrPheSerGlyAlaLeu 120
Db 301 GACACCTTCCCAAGGGGAAAGCTGCAGCGATGGCGAGCTACATCAGCGCGGCTTG 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 CGGGCGACGCGCTTGGCAATGGCGCTGCTTCAGTACATGACGCTCGGATCAACAAG 420
QY 141 GlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 421 GCGGATGGCTGCGCGCTCCGTCAAACCGCTGACCCGCTCATTTCCGGAGCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr 180
Db 481 GCGGCCATGGACCAAGTGGGACCAAGATGATGACCGCGAGGCTGATTCACATTAC 540
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaLaseValLysArgHisSer 200
Db 541 CTGAGGCGCTCGCGCGGACAGCTCCACAGATGCGATGGCCGCTCGGGAAGCGCACTCG 600
QY 201 ProSerLeuAlaArgGlnValIleAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
Db 601 CCAAGCCTTGCTCGACAGGTTCTGGACACGCGGGGTTGCGGTTTCAGAGCTCGCGCGCG 660
QY 221 AsnAlaValArgThrValIleAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 661 AACGCGGTACGTCACCGATTTGGCTCCGGCACTGGCTCCAGACCCGCGTGCAGGGTGTCT 720
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 721 GTGACCTTGATGATGATGAGGGGGGTCTGGCTGCGCAACGACGCTTGGCAACCGC 780
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 781 CTGCTAGTGTGACGTGCGCTGATTCACACACGCGGGGCTCATTAAGTGTGCTTGGAG 840
```

QY	281	Aspysgluprolysalaglinsleuserruglinsuasrptpleuglnalatrlytslaa	300
Db	841	GATMAAAGCCCAAGGCTCACTAGACGAAGAAGACGTGCTCGAGCTTATTAACA	900
QY	301	IlleysseralaserTYrSerGlyAlaAlaLeuasnAlaGlyLysArgMetAlaGlyLeu	320
Db	901	ATCAAAATCGGCCAGACTACTCGGGTGGCGGCTCAACGCTGGCAAGGGATGGCCGGTCTG	960
QY	321	ProleuaspmetalathraspalametGlyAlaValArgSerLeuValSerAlaSerSer	340
Db	961	CCACTGGATATGGGACCGACCGCAAGGGGTGGGTAAAGAGCTGTGTACACCGTCCAGC	1020
QY	341	LeuthrGlnasnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuInGlu	360
Db	1021	CTGACCCAAAGCGCTCTGGGCGCTGGCGGGTGGCTTTGCAAGGGGTAGCAAGTTGCAAGAG	1080
QY	361	MetAlaThrLysasnLierthrasProAlaThrLysAlaAlaValSerGlnLeuThrAsn	380
Db	1081	ATGGCGACGAAAAATATCACCGACCGGGGACCAAGCGCGGTCAAGTACGTTGACCAAC	1140
QY	381	LeuAlaGlySerAlaAlaValPheAlaGlyTYrPThrAlaAlaLeuThrThrasPro	400
Db	1141	CTGGCAGGTTTCGGCAGCGCGTTTCGCAGGCTGGACCAAGCGCGCTGCACAACGATCC	1200
QY	401	AlaAlaLysLysAlaGlySerPheIleGlnAspThrValLysSerThralaSerSerThr	420
Db	1201	GGGCTGAAAAAGCCGCGGTGCTTCATACAGGACCGGTAAATCGACTGCATCCAGTACC	1260
QY	421	ThrcGlyTYrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetLysGly	440
Db	1261	ACAGGCTACGTAAGCGGCACGACCGTCAAACTGGCCAAACCGCTCAAAACATGGGCGG	1320
QY	441	GluAlaIleThrIsthrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg	460
Db	1321	GAGCGCATCACCCATACCGGCGCCACGCTTGCCGAATACGCTCAATACCTGCGTCAACGC	1380
QY	461	ProhlaArgGluAlaAspIleGlnGlyGlyTYrThAlaAlaSerProserGluIlePro	480
Db	1381	CGGCTCGTGAAGTGTATTAAGAAGGGGGGCACGGCGCTTCTCCAAGTAAATATACCG	1440
QY	481	PheArgProMetArgSer 486	
Db	1441	TTTTGGCCTATGGCGTGC 1458	
RESULT 3			
ID	AAD20405/c		
AC	AAD20405 standard; DNA; 30365 BP.		
XX	AAD20405;		
XX	03-JAN-2002 (first entry)		
DE	P. syringae pv. tomato (Pto) DC3000 Conserved Effector Loci (CEL) DNA.		
KW	Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;		
KW	Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;		
KW	eukaryotic cell death; cancer; ds.		
OS	Pseudomonas syringae.		
XX			
PN	WO200175066-A2.		
XX	11-OCT-2001.		
XX			
PF	03-APR-2001; 2001WO-US10698.		
XX			
PR	03-APR-2000; 2000US-194160P.		
PR	11-AUG-2000; 2000US-224604P.		
PR	17-NOV-2000; 2000US-249548P.		
XX			
PA	(CORR) CORNELL RES FOUND INC.		
PA	(UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.		
PA	(UYNE-) UNIV NEBRASKA.		

[illegible]

|||||
Db 23624 CCGGCGACGCCCTTTGCCATGGCTGCTGCTTCACTACATGACGCTCGCATCAACAG 23565
QY 141 GYAAsPTTReuProAlaProLeuLysProLeuThrProLeuLLeSerGlyAlaLeuSer 160
Db 23564 GCGGATTTGGCTCGCGCTCGCTCAAAACCGCTGACCCCTCATTTCCGAGCGCTGTGC 23505
QY 161 GLAAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuH1sTYR 180
Db 23504 GCGGCGATGACAGGCTGGGACACCAAGATGAGACCGCGGCGGTGATCTGCATTAC 23445
QY 181 LeuSerAlaSerProAspArgLeuH1sAspAlaMetAlaLeuSerValLysArgH1sSer 200
Db 23444 CTGAGCGCGCTCGCGGACAGGCTCCACGATGCGATGCGCTTGGTGGAAGCCCACTCG 23385
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrYrSerAlaArg 220
Db 23384 CCAAGCCTTGCTCGACAGGTTCTGGACACGGGGGTTGCGGTTCAGACGTACTCGCGCGC 23325
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 23324 AAGCGCGTACGTACCTATTGGCTCCGCACTGGCTCCAGACCCCGCTGCAGGGTGTCT 23265
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 23264 GTGGACCTTGCTGTATCATGATGGCGGGGTGCTGCTGCCACCGAGCTTTGGCAACCCG 23205
QY 261 LeuLeuSerValGlnSerArgAspH1sGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 23204 CTGCTCAGTGTGCTGACGCGGTGATCACACGCGTGGCGGTGATTAAGCTTGAAG 23145
QY 281 AspLysGluProLysAlaGlnLeuSerGlnGluAsnAspThrLeuLysArgLysAla 300
Db 23144 GATTAAGACCCCAAGGCTCACTGACGACGAGAAGAACGAGCTGGCTCAAGGCTTAAAGCA 23085
QY 301 IleLysSerAlaSerTyriserGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 23084 ATCAATCGGCGCAGTACTCGGGTGGCGGCTCAACGCTGCAAGCGGATGGCGGCTCTG 23025
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 23024 CCACTGGATATGGCGACCGACCGCAATGGGTGGCTGAAGACCTGTGTGCAGCGTCCAGC 22965
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
Db 22964 CTGACCCAAACGCTGTGGCCCTGGCGGGGTGCTTGCAGGGGTAGGCAAGTTGCAAGAG 22905
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
Db 22904 ATGGCGACAAAATATACACCGACCGGCGACCAAGCGGGTCAAGTCAGTTGACCAAC 22845
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrAlaAlaLeuThrThrAspPro 400
Db 22844 CTGGCGGGTTCGGCACCGCTTTTTCGAGGCTGGACACCGCGCCGCTGCACACCGATCCC 22785
QY 401 AlaValLysLysAlaGlnSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 22784 GCGGTAAAAAGCCAGTCGTTTCATACAGACACCGGTGAATCGCTGCATTCATAC 22725
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 22724 ACAGGCTAGTACCGCACACGACCGTCAAACTGGCAAGCCGTCAAAGCATGGCGGG 22665
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAspThrValAsnAsnLeuArgGlnArg 460
Db 22664 GAGGCGATACCATATACCGCGCGCTTCCGCAATACGCTGCTCAACCG 22605
QY 461 ProAlaArgGlnAlaAspIleGlnGlnGlyGlyThrAlaAlaSerProserGlnIlePro 480
Db 22604 CCGGCTCGTGAAGCTATATAGAAAGAGGGGCGACGCGGCTTCTCAAGTCAATACCG 22545
QY 481 PheArgProMetArgSer 486
|||||

Db 22544 TTTGGGCTATGCGGTGCG 22527
RESULT 4
AAD20438
ID AAD20438 standard; DNA; 1464 BP.
XX
AC AAD20438;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 DNA.
XX
KW Conserved Effector Loci: CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci: EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer; ds.
XX
OS Pseudomonas syringae.
XX
FH Key location/Qualifiers
FT 1.1464
FT /tag= a
FT /product= "P. syringae pv. tomato HopPtoA2 protein"
XX
PD MO200175066-A2.
XX
XX 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
XX
XX (UYNE-) UNIV NEBRASKA.
XX
PI COLLIER A, Alfano JR, Charkowski AO:
XX
XX WPI: 2001-639361/73.
XX P-PSDB: AAE12603.
XX
XX New nucleic acid molecules encoding proteins or polypeptides of
XX Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
XX genomic sequences, for imparting disease resistance to plants -
XX
XX Claim 1; Page 64; 217pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule comprising a
XX nucleotide sequence encoding proteins or polypeptides of Pseudomonas
XX conserved effector loci (CEL) and Exchangeable Effector Loci (EEL)
XX genomic sequences. CEL and EEL DNA are useful for imparting disease
XX resistance to a plant, by transforming a plant cell with the nucleic acid
XX and regenerating a transgenic plant from the transformed plant cell,
XX where the transgenic plant expresses a heterologous DNA molecule under
XX conditions effective to impart disease resistance, or by treating a plant
XX with an isolated protein or polypeptide, by applying the protein or
XX polypeptide in an isolated form or by applying a non-pathogenic bacteria
XX which secretes the protein or polypeptide, under conditions effective to
XX impart disease resistance to the treated plant. CEL and EEL proteins
XX are useful for causing eukaryotic cell death, by introducing a cytotoxic
XX pseudomonas protein into a eukaryotic cell under conditions effective to
XX cause cell death. CEL and EEL proteins are also useful for treating a
XX cancerous condition, by introducing a cytotoxic Pseudomonas protein into
XX cancer cells of a patient under conditions effective to cause death of
XX cancer cells, and thus treating the cancerous condition. The method
XX further involves administering a targeted DNA delivery system
XX comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
XX to the patient, where the targeted DNA delivery system delivers the
XX DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
XX is expressed in the cancer cells. The present sequence is a DNA
XX encoding Pseudomonas syringae pv. syringae HopPysA homolog protein.

SQ Sequence 1464 BP; 325 A; 392 C; 442 G; 305 T; 0 other;

Alignment Scores:

Pred. No.:	4.27e-129	Length:	1464
------------	-----------	---------	------

Score:	1869.50	Matches:	381
Percent similarity:	85.19%	Conservative:	33

Percent Similarity:	88.1%	Conservative:	55
Best Local Similarity:	78.40%	Mismatches:	71

Query Match:	77.51%	Indels:	1
--------------	--------	---------	---

DB: 22 Gaps: 1

US-09-825-414-7 (1-486) X AAD20438 (1-1464)

US-09-825-414-7 (1-486) X AAD20438 (1-1464)

QY	1	MethistileksnaIdgVdgvValgIngnInProProValThralaThrAspSerPheArgThr	20
Db	1	ATGCACATCAACCAATCCGCCGCAACCAACCGCGCTGGCTTCGATGAGAGATTTCGACCA	60
QY	21	AlaserAspAlaserLeuAlaserSerSerValArgSerValSerSerAspGlnArg	40
Db	61	GCCTCCGACGGCTCCCTGGCTTCGAGTTCTGTGGGAGTGTGCACACCTACCTCGTCCGC	120
QY	41	GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisIstysLeuPro	60
Db	121	GATCTACAGGTATTACCGATTATCTGAACATCAAGCTGTTCGCTGGCGACAGTTTTCG	180
QY	61	ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThAla	80
Db	181	GTAATTAGGCTACCCGAGGTAGACCGTGAAGCCGCTTTCACACAAACGACAGATGATGCG	240
QY	81	LeuIleGluThrArgAlaIaserArgLeuHisPheGluGluThrProAlaThrIleAla	100
Db	241	TTGGTAGAGACACCGCCACACCGCTGTACTCCGAAAGGGAGACCCCCGACCAATCGCC	300
QY	101	AspThrPheAlaValAspAlaGlyIstysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu	120
Db	301	GAACATTCGCCGCAAGGGGAAAGATTGTGACCGTTTGGCGACGACCGGATCAAGTCTTT	360
QY	121	ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnIleThrMetGlnProAlaIleAsnLys	140
Db	361	GAGAACACCGCATTTGCCCGCTGCTCGCTTCAGTACATGACAGCTCGGACATCAACAG	420
QY	141	GlyAspThrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160
Db	421	GGCGATTGGCTAGACACGCGCGCTAACCCGCTGACCCCGCTCATTTCCGAGCGCTTCG	480
QY	161	GlyAlaMetAspArgGlnValGlyThrLysMetMetSerPargAlaThrGlyAspLeuHisTyr	180
Db	481	GGAGCCATGGAGACAGAGTGGGAGCCAAATATGATGATGTCTGCGAGGGGTATCTGCATTAC	540
QY	181	LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer	200
Db	541	CTGAGCACTTCGCCGCGACAAATTGCATGATGTGCATGCGCCGATGCGGAAGCGCCACTCG	600
QY	201	ProSerIleuAlaArgGlnValLeuAspThrGlyAlaAlaValGlnThrTyrSerAlaArg	220
Db	601	CTCGCGCTGGCTGCACAGGTGTGTGCAATGGGGGATTCCAGTGCACAGAGTTCTCGGCGCTA	660
QY	221	AsnAlaValArgThrValLeuAlaIleProAlaIleuAlaIaserArgProAlaValGlnGlyAla	240
Db	661	AATGTGGTGGCTACCGATTGGTGGCTCCAGCACTACCGTCCAGACCGCTGGGTGCAGCGTCT	720
QY	241	ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg	260
Db	721	GTTGATTGTTGGCTGATCTACAGCGCGGTGGCTTGGTTCGCAATGCAAGGCTTTGGGCGACCG	780
QY	261	LeuLeuSerValGlnSerArgAspHisGlnArgGlyLysAlaLeuValIleuGlyLeuLys	280
Db	781	ATGCTCAGTGTGCATCGACCGCGGATCAACCTGCGTGGGGGCGCATTCGTACTTGGCATCAAA	840
QY	281	AspLysGluProLysAlaGlnIleuSerGlnGlyAsnAspThrLeuGlnAlaTyrLysAla	300
Db	841	GATTAAGAGCCCAAGCGCGCTTAAAGTGAAGAAACTATTTGGCTTGAAGCTTCAAAAGCG	900
QY	301	IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320

Db	901	ATCAAGTGGCCACTACTCACTGAGTGGCGCGCTCAATCGGGCAAGCGGATGGCGGCTCTG	960
Qy	321	ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer	340
Db	961	CCACTGGACGTCGCCGACCGACGGCTCAAGCGGTGGAAAGCTGTGGTGTGGCCACCACG	102
Qy	341	LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyAlaGlyLeuGlnGlu	360
Db	1021	CTGACAAAAAATFGGCGTGGCCCTAGCCGGTGTACGCCGGGGTAAGTAAATTGCAGAAA	108
Qy	361	MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn	380
Db	1081	ATGGCAGCAAAAAATATCACTGATTCCGGCCGACCAAGGCTCGGTATGATCAGCTGACAC	114
Qy	381	LeuAlaGlySerAlaAlaValPheAlaGlyThrPheThrAlaLeuThrThrAspPro	400
Db	1141	CTGGTGGGTGGTGAAGCGTTTGGCAGCGTGGACACCCTCGTGACTGGCGACTGACCT	120
Qy	401	AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr	420
Db	1201	GCGGTAAAGAAAGCCGAGTCTTTATATACAGATTAAGTGAATCGAACCGCATCTAGTACC	126
Qy	421	ThrGlyThrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly	440
Db	1261	ACAACATGTATGTGGCCGACCAACCCGTCAAACTGGCGAAAAACAGTCAAGCATGACGGG	132
Qy	441	GluAlaIleThrHisThrGlyAlaSerLeuArgAspThrValAsnLeuArgGlnArg	460
Db	1321	GAGGGGATCTCCAGACCGGTCCAGCTTACGCAAGTACTGTCAATTAACCTGGCTCATCGC	138
Qy	461	ProAlaArgGluAlaAspIleGluGlyGlyThrAlaAla--SerProSerGluIle	479
Db	1381	TCCGTCGGGAAGCTGGAATTCGAAGAAGTGGGATTTCGGCGTTTCTCGAAGTGAACA	144
Qy	480	ProPheArgProMetArg	485
Db	1441	CCGTTTCAGCTCAGGCGT	1458
RESULT 5			
AA199683/c			
ID	AA199683 standard; DNA; 4403765 BP.		
XX	AA199683;		
AC			
XX			
DT	15-JAN-2002 (first entry)		
XX			
DE	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.		
XX			
KW	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;		
XX	variation; epidemiology; patient treatment; epidemic monitoring; ds.		
OS	Mycobacterium tuberculosis.		
XX			
PN	US6294328-B1.		
XX			
PD	25-SEP-2001.		
XX			
PF	24-JUN-1998; 98US-0103840.		
XX			
PR	24-JUN-1998; 98US-0103840.		
XX			
PA	(GENO-) INST GENOMIC RES.		
XX			
PI	Fleischmann RD, White OR, Fraser CM, Venter JC;		
XX			
DR	WPI; 2001-647261/74.		
XX			
PT	Evaluating strain variation of Mycobacterium tuberculosis, comprises		
XX	determining the nucleotide sequence of the strain at positions in the		
PT	genome corresponding to positions where M. tuberculosis strains CDC		
XX	1551 and H37Rv differ -		

PS Claim 2, Fig 63; 257bp; English.

XX The present invention relates to a new isolated polynucleotide comprising a sequence present in *Lactobacillus rhamnosus* strain HN01 and encoding a polypeptide capable of modifying the flavour, aroma, texture, nutritional and health benefits of milk-derived products, and/or survivability of microbes in dairy manufacturing processes. The polynucleotides are useful for improving the properties of microbes used in the manufacture of milk-derived products such as cheeses, yogurt, fermented milk products, sour milks and buttermilk; in modifying the flavour, aroma, texture and health -related benefits of milk-derived products and in increasing the survival of microbes during industrial fermentation processes. The bacteria may be used to increase resistance to enteric pathogens and anti-infection activity, including treatment of rotavirus infection and infantile diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis; liver cancer reduction; reduction of small bowel bacterial overgrowth; CC immune system modulation and treatment of autoimmune disorders and CC allergies; treatment of allergic responses to foods; reduction of blood CC lipids and prevention of heart disease; antihypertensive effect; CC prevention and treatment of urogenital infections, *Helicobacter pylori*, CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and CC irritable bowel syndrome; modulation of endocarditis; and for improved CC protein and carbohydrate utilization and conversion. The transgenic CC microbial population can be administered to a mammal as an anti-carcinogenic agent. The present sequence is *Lactobacillus rhamnosus* outer membrane protein rompa gene.

XX Sequence 4645 BP; 1279 A; 1062 C; 1261 G; 1043 T; 0 other;

Alignment Scores:

Pred. No.:	0.186	Length:	4645
Score:	148.50	Matches:	126
Percent Similarity:	36.18%	Conservative:	69
Best Local Similarity:	23.38%	Mismatches:	233
Query Match:	6.16%	Indels:	112
DB:	24	Gaps:	21

US-09-825-414-7 (1-486) x AAD31881 (1-4645)

```
QY 13 ThrAlaThrAspSerPheAla-----SerAspAlaSerLeuAlaSerSer--- 29
    |||||
DB 2864 ACTGCAACATCTGCGTATGTAAGTGTGCACAGTTCCAGTGCACGTGAAGCGGTTCAAT 2923
    |||||

QY 30 -----SerValArgSerValSerSerPheGlnGlnArgGluIleAsnAlaIleAlaAsp 47
    |||
DB 2924 GCTATCAGGAGAGGCTCCAGGCGCCAGTACGCTGCGTCCGCGCAGTGCAGCCAA 2983
    |||

QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProPro-----AlaAspSer 64
    |||
DB 2984 CATGCCAGACCGCTGCGAGTGGCGCATCCAGTATCCGAAGATAGTGGGATTTCAGTCA 3043
    |||

QY 65 AlaAspGluGlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThr 84
    |||||
DB 3044 CTAGCCAGTACGCTGCGAGGAGGACCAAGGCAAGCAGTAAACGCGAGTGCAGCAACC 3103
    |||||

QY 85 ArgAlaSerArgIleHisPheGluGluGluThrProAlaThr---IleAlaAspThrPhe 103
    |||||
DB 3104 AGCCCGCGCGCGTGTGTTTCAAGTCTGCCAGTATGATGACAGGCGCAAGAGCGCT 3163
    |||||

QY 104 AlaLysAlaGluLysLeuAspArgLeuAlaThrThr-----SerGly 118
    |||
DB 3164 GCAAGTCCGATGTGTGGCAAGCAGTGGCGCCAGCAGGCTAACAGTAATGCCAGTGGC 3223
    |||||

QY 119 AlaLeuArgAlaThrPro-----PheAlaMetAlaSer 129
    |||
DB 3224 GCACCCAGTGGGACCAAGGCTGTGATAGCAAAAGCCGACGAGGATTCGACAGTGCAGCG 3283
    |||||

QY 130 -LeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspThrLeuProAla----- 146
    |||||
DB 3284 AGTCTGTGACGACAGCATGTCGAAGGTGTCAGAG-----CAGTTGCCAGGGAAGCGCG 3337
    |||||

QY 147 ----ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla-----Me 163
    |||||
```

```
DB 3338 AGTCCCGCGGACATCCGATGACTCG-GTAGCTTTAGTCCCGCAGTCCGCGTGCAGGCTT 3396
    |||||
QY 163 TAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAl 183
    |||||
DB 3397 TGACAAAGCTCCAGCGCTGCGGAAGCGGAGCTTCA-----AGTGC 3438
    |||||

QY 183 AserProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLe 203
    |||
DB 3439 CGCAGCGCGGCTCTAGTTACGGCGGACGTCACAGCAGCAGCAGAGTGGCGCAACTCCAG 3498
    |||||

QY 203 uAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTr----- 217
    |||
DB 3499 TGCACGGAACCGGGTCAAGCATCAACCGCAACATCTGTATGCTAGTGCCTCAAGTTC 3558
    |||||

QY 218 -----SerAlaArgAsnAlaVal 223
    |||||
DB 3559 GAGTGCAGTGAACCGGTTCAATGCTCATCAGCAGCGCTCCAGTCCAGTGAAGGCAC 3618
    |||||

QY 223 L-----ArgThrValLeuAlaProAlaLeuAlaSerArgPr 235
    |||||
DB 3619 TGGCCATGCAAGTATGCTGTACAACTCAAGCAGTCCGATCCAGTGTGCTCCAGGTA 3678
    |||||

QY 235 oAlaValGlnGlyAlaValAspLeuGlyValSerMetAlaGlyLysLeuAlaAlaAsnAl 255
    |||||
DB 3679 CCCAAGTATATGTTGGATCCAGTCAAGATGTAATGCGTCCAGTGCAGCAAGTACTGCG 3738
    |||||

QY 255 aGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLe 275
    |||
DB 3739 ATCC-----AGTGGCGGTATGTCGCGCACAAAGTGAAGCGCTTCAGCGC 3780
    |||||

QY 275 uValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluLysAsnAspTrpLe 295
    |||||
DB 3781 ATGTCGGCTGCAAGTATGCTATGTAACAAGCAAGATATGCTTCCAGTGGAGTGTGTT 3840
    |||||

QY 295 uGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLys 315
    |||||
DB 3841 ATCAAGCAGTGTCTGCGAGTGTGCGTCCAGCGCGGCGCAGTCCGATCCAGTGTGCAAA 3900
    |||||

QY 315 s-----ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAl 331
    |||||
DB 3901 GCGTGTAAACAGTATGCTGTGGGTATATATCTCATGACAGCAAGTCCAGTGTGCAAGCAG 3960
    |||||

QY 331 aValArgSerLeuValSerLeuAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyL 351
    |||||
DB 3961 TGCTAAAGCCCTCAAAAGTCAAGCAGAGTGGCCGACAGTGTCTCTCT----- 4009
    |||||

QY 351 yPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThr 371
    |||
DB 4010 -----TCTGATGATTCGGTACGCTTCCAGCGTGC 4038
    |||||

QY 371 rLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTr 391
    |||||
DB 4039 CAGTGCCTGTTGCTGTACGATGCTAAGGCAAGTGAAGCGCGCATGTAGATCCAGCGC 4098
    |||||

QY 391 pThrThrAlaAlaLeuThrThrAspProAlaVal-----LysLys 404
    |||||
DB 4099 TACACTGTGCTCATTTGTTCCGCCACATCTTGGCTGTATCAGAGTCCACAGGCTCAAC 4158
    |||||

QY 404 sAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTrVal 424
    |||||
DB 4159 CGGTGGCTCCCATAT-----TTGCCAAGTACGTGGGAGACAGACAGCATAGTATAT 4209
    |||||

QY 424 l-----AlaAspGlnThr-----ValLysLeuAlaLysThrVa 435
    |||||
DB 4210 ACCATCGGTCAGACGCCAACACAGACAGCAAGCCAAACAGACAGCAACACAAACGAA 4269
    |||||

QY 435 lLysAspMetGlyLysGlnAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAs 455
    |||||
DB 4270 GCCAACAACAAGCCGCTAAACA---ACCCAGACAGGT---TCATTACCGCAAAAGGATCA 4323
    |||||

QY 455 nasnLeuArgGlnArgProAlaArgGluAlaAspIleGluGluGlyLysThrAla 473
    |||||
DB 4324 TGCAGGAGGCAATATGCTACCGCAGACCGGTGATGATGATGATGAAGCGGTACTTCT 4378
    |||||
```

RESULT 8	
AAQ29471	
ID	AAQ29471 standard; DNA; 6744 BP.
XX	
AC	AAQ29471;
XX	
DT	04-MAR-1993 (first entry)
XX	
DE	Extracellular factor related protein gene.
XX	
XX	EF*, detection; prevention; screening; diagnostic; ss.
OS	Streptococcus suis type II (non-pathogenic).
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	361..5826
FT	/*tag= a
FT	361..498
FT	/*tag= b
FT	499..5826
FT	/*tag= c
FT	66..71
FT	/*tag= d
FT	/*note= "-35 region"
FT	89..94
FT	/*tag= e
FT	/*note= "-10 region"
FT	153..158
FT	/*tag= f
FT	/*note= "-35 region"
FT	176..181
FT	/*tag= g
FT	/*note= "-10 region"
FT	350..356
FT	/*tag= h
FT	/*note= "ribosome binding site"
FT	2869
FT	/*tag= i
FT	/*note= "start of repetitive unit R1"
FT	3097
FT	/*tag= j
FT	/*note= "start of repetitive unit R2"
FT	3292
FT	/*tag= k
FT	/*note= "start of repetitive unit R3"
FT	3520
FT	/*tag= l
FT	/*note= "start of repetitive unit R4"
FT	4087
FT	/*tag= m
FT	/*note= "start of repetitive unit R5"
FT	4381
FT	/*tag= n
FT	/*note= "start of repetitive unit R6"
FT	4609
FT	/*tag= o
FT	/*note= "start of repetitive unit R7"
FT	4837
FT	/*tag= p
FT	/*note= "start of repetitive unit R8"
FT	5065
FT	/*tag= q
FT	/*note= "start of repetitive unit R9"
FT	5293
FT	/*tag= r
FT	/*note= "start of repetitive unit R10"
FT	5521
FT	/*tag= s
FT	/*note= "start of repetitive unit R11"
FT	6554..6566
FT	/*tag= t
FT	/*note= "dyad symmetry region 1a"

FT	misc_feature	6571..6583	/*tag= u	/note= "dyad symmetry region 1b"
FT	misc_feature	6611..6625	/*tag= v	/note= "dyad symmetry region 2a"
FT	misc_feature	6631..6644	/*tag= w	/note= "dyad symmetry region 2b"
FT				
PN	W09216630-A.			
PD	01-OCT-1992.			
PF	19-MAR-1992;	92WO-NL00054.		
PR	21-MAR-1991;	91NL-0000510.		
PA	(DIER-) CENT DIERGENESKUNDIG INST.			
PI	Smith HE, Vecht U;			
DR	WPI: 1992-349215/42.			
DR	P-PSDB: AAR27745.			
PT	Deoxyribonucleic acid encoding virulence characteristic of			
PT	Streptococcus suis - useful for antibody and polypeptide for			
PT	diagnosing and preventing infections in pigs and humans			
PS	Claim 2; Fig 1b; 86pp; English.			
XX				
CC	The sequence is that of the extracellular factor related protein gene			
CC	from Streptococcus suis type II (non-pathogenic) which allows the			
CC	detection and the prevention of infections by S. suis in a more			
CC	effective manner than was previously possible. It facilitates screening			
CC	of e.g. pigs and elimination of infected and carrier pigs can then be			
CC	carried out. The new diagnostic tests can distinguish between			
CC	avirulent and virulent strains. It may be used in the prodn. of			
CC	a vaccine. See also AAQ29470 and AAQ29472.			
XX				
SO	Sequence 6744 BP; 2402 A; 1231 C; 1535 G; 1576 T; 0 other;			
Alignment Scores:				
Pred. No.:	0.569	Length:	6744	
Score:	144.50	Matches:	109	
Percent Similarity:	37.04%	Conservative:	81	
Best Local Similarity:	21.25%	Mismatches:	208	
Query Match:	5.99%	Indels:	115	
DB:	13	Gaps:	22	
US-09-825-414-7 (1-486) x AAQ29471 (1-6744)				
QY	AlaThrAspSerPheargThrIaSerAspAlaSerLeuAlaSerSerValArgSer 33			
DB	4024	GCTAGAGATGCGATTGAATGGCTAAAGATTAAGAATTGGTAAGGACCAATCCGACACA	4083	
QY	34	ValSerSeraspGlnIaargGluIleasnAlaIleAlaAsp-----	47	
DB	4084	GAAGAAGAAGAGCTACTTAATAATGATAGAGAACAATTGCAGAAAGATACGGCAAACTATC	4143	
QY	48	-----TyrLeuThrAsp 51		
DB	4144	GAGGACAAATCCAAACTTGTCCAGATGAGATTAAGCAAGCGGAATTAAAAAGCTAACTGAC	4203	
QY	52	HisValPheIaIaIaHisIysLeuProProIaIaAspSerAlaAsp-----	66	
DB	4204	GCTGTG---GCAAAAACCTTTAGCAACCATGCGTGCACAAATGACATGACGCGTACGCAAGA	4260	
QY	67	-----GlyGlnAlaIaIaValAspValHisAsnAlaGlnIleThrAlaLeuIleGlu 83		
DB	4261	GCAGAAAAAGCTCAGACCCCTACGACATCTTGGAAAAAGCTAAAGAAACACAGAAAAATTGCA	4320	
QY	84	ThrArgAlaIaSer-----ArgLeuHisPhe-----GluGlyGluThrProIaIaThr 98		

D	4321	GATTAACCTGCGATTGATAGCTTGA	CTATACCTTGCAAGATGCTGACCTTGA	AGCTACT	4380
Q	99	IIeAlaAspThrPheAlaIuValGlu	LeuAspArgLeuAlaThrThrSer	Gly	118
D	4381	AAACACAGAT-----GCTAAGACC	---AAGATTGCTAAAGATGACGCGCTG	CTAAAGA	4431
Q	119	AlaLeuArgAlaThrProPheAlaMet	AlaSerLeuGlnTyrMetGlnPro	AlaIle	138
D	4432	GCAATTCGCAACATCCA-----			4449
Q	139	AsnLysGlyAspThrPheProAlaPro	LeuLysProLeuThrProLeuIle	SerGly	158
D	4450	AACCTTGACACAC-----GCAGAGAA	AAAAAACCCTTCACCGATGCGGTAA	ATGCAGAA	4500
Q	159	LeuSerGlyAlaMetAspGlnValGly	ThrLysMetMetAspArgAlaThr	GlySerLeu	178
D	4501	GTACCCAAACCTAACGACGAATTTCAG	CT-----GCACACAGC-----		4539
Q	179	HisTyrLeuSerAlaSerProAspArg	LeuHisAspAlaMetAlaSerVal	LysArg	198
D	4540	-----CCACAGATGTTCTAA	AAAAAGAAAGAGATGACGATGTCG	ACCATTCAGAA	4590
Q	199	HisSerProSerLeuAlaArgGlnVal	LeuAspThrGlyValAlaValGln	ThrTyrSer	218
D	4591	GATGTTCTGACGACGCTTAAACCAAG	TGCTAAGAAATACATTGCTTAAAGAT	GCACCGCT	4650
Q	219	AlaArgAsnAlaValArgThrVal	LeuAlaProAlaLeuAlaSerArg	ProAlaValGln	238
D	4651	GCTTAAACACCAATTTGGCTTCAT	CCAAACTTGACAGCGCAGAGAA	AAAACTTCCAC	4710
Q	239	GlyAlaValAspLeuGlnValSerMet	AlaGlyGlyLeuAlaAlaAsnAla	GlyPheGly	258
D	4711	GATCGGTGATGATCAGAGATGCAAG	AAAGCTAACGACCAATTTGACGTGCA	-----	4761
Q	259	AsnArgLeuLeuSerValGlnSerAla	GspHisGlnArgGlyValAlaLeu	-----	275
D	4762	ACCAGCCCAACAGATGTTCTAA	AAAAAGAGAGATGCAGGTTCAGCG	CAATTCACAGAT	4821
Q	276	ValLeuGlyLeuLysAspLysGlnPro	LysAlaGlnLeuSerGlnGlnAsn	AspThrLeu	295
D	4822	GTTCTTACGACACTAAACAAGTGT	TAAGTAATACATTGCTTAAAGAT	CTCCAC-----	4875
Q	296	GluAlaTyrLysAlaIleLysSer	Ala-----SerTyrSerGlyVal	Ala	310
D	4876	-----GCTGCTAAGTCAGCCAT	TGACCGCAATCCAAACTTGACAGAT	GCACAG	4923
Q	311	LeuAsnAlaGlyLysArgMetAlaGly	LeuProLeuAspMetAlaThrAsp	AlaMetGly	330
D	4924	AAGCAATCACCTTAAGAAAGCAG	TGTGCTGATGCTTAAAGCTGC	ACACAGATGCATTAAT	4983
Q	331	AlaValArgSerLeuValSerAlaSer	LeuThrGlnAsnGlyLeuAlaLeu	AlaGly	350
D	4984	GCTTCACCAAGTCCATGCTCAAG	CGCAATGCGCAGAGACAA-----		5025
Q	351	GlyPheAlaGlyValGlyLysLeu	GlnGluMetAlaThrLysAsnIle	ThrAspPro	370
D	5026	-----GGCGTAGGTTCAATC-----			5064
Q	371	ThrLysAlaIaIaIaSerGlnLeu	ThrAsnLeuAlaGlySerAlaAla	ValPheAlaGly	390
D	5065	AAACCAAGATCTTAAGAACAGAT	TGCCAAAGAAAGTTGCCGCGAGCT	-----	5109
Q	391	TyrThrThrAlaIaLeuThrThr	AspProAlaValLysLysAlaGlu	-----SerPheIle	409
D	5110	-----AAACAACCAATTGATGCC	CAATCCGAACCTTATCAGATGC	ACAGAGAAAGACCTTCT	5163
Q	410	GlnAspThrValLysSerThrAlaSer	SerThrThrGlyTyrValAlaAsp	IleThrVal	429
D	5164	AAGAAACGCGTAGTGCAGATGCT	TAAAGCTACGACGATGCATTAAT	GATGCTTCAACAGT	5223
Q	430	LysLeu---AlaLysThrValLys	AspMetGlyGlyAlaAlaIle	ThrHisThrGlyAla	448

Db	5224	CCATCGACGACGCAATCGCAGAGGACAAAGCG-----GTAGTCA	5265
Qy	449	SeLeuAaRgsnThrValAsnAsnLeuArgIn-----ArgProAlaArgGlu	464
Db	5266	ATGCCCAAGATGTTCTTGACGCGAGCAACAAGATGCTAAGATTAAGATTGCTAAAGAA	5325
Qy	465	AlAspIleGluGluGlyGlyThrAlaIleSerProSer	477
Db	5326	TCCGACGCTGCTAAGTCAGCCATTGACGGAATCCAAAC	5364
RESULT 9			
ID	AAAl0263	standard; DNA; 11893 BP.	
XX	AAAl0263:		
XX	03-JUL-2000	(first entry)	
DE	Bordetella pertussis filamentous haemagglutinin gene, fhaB.		
XX			
KW	Filamentous haemagglutinin; FHA; fhaB gene; surface-associated;		
KW	cellular adhesion; antigenic; immunogenic; whooping cough; pertussis;		
KW	diagnosis; prophylaxis; vaccine; ds.		
XX			
OS	Bordetella pertussis.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	253..11043	
FT		/tag= a	
FT		/product= "B. pertussis filamentous haemagglutinin"	
XX			
PN	US6036960-A.		
PD	14-MAR-2000.		
XX			
PF	01-SEP-1994:	94US-0299941.	
XX			
PR	13-NOV-1989:	89US-0436297.	
PR	10-AUG-1992:	92US-0928619.	
PR	27-OCT-1988:	88US-0263648.	
XX			
PA	(RELW/) RELMAN D. A.		
PA	(DOME/) DOMENIGHINI M.		
PA	(RAPP/) RAPPUOLI R.		
PA	(FALK/) FALKOW S.		
PI	Falkow S, Domenighini M, Relman DA, Rappuoli R;		
DR	WPI: 2000-269872/23.		
DR	P-PSDB: AAY87407.		
XX			
PT	New isolated nucleic acid encoding the filamentous hemagglutinin of		
PT	Bordetella pertussis, useful for diagnosis, treatment and prevention of		
PT	whooping cough -		
XX			
PS	Claim 3; Columns 11-28; 17pp; English.		
XX			
CC	This sequence represents the filamentous haemagglutinin gene, fhaB, from		
CC	Bordetella pertussis. Bordetella pertussis is a small Gram negative		
CC	bacillus found only in humans and which causes whooping cough (pertussis)		
CC	in children. Filamentous haemagglutinin (FHA) is a 368 kD surface-		
CC	associated protein which is thought to be one of the most important		
CC	factors in mediating the adhesion of the bacterium to a eukaryotic		
CC	(human) cell. FHA also stimulates an immune response in humans following		
CC	disease, and has been shown to act as an immunoprotective antigen in an		
CC	animal model. The FHA gene, or its expression products, are used for		
CC	prevention or treatment of pertussis, particularly in vaccines. It may		
CC	also be used to increase the amount of FHA in live or dead Bordetella		
CC	pertussis or other organisms. Fragments of the gene may be used as		
CC	diagnostic hybridisation probes or as antisense modulators to reduce the		
CC	infectivity of Bordetella pertussis. FHA, or fragments thereof, may be		
CC	used in immunoassays to detect anti-FHA antibodies, and to raise		
CC	antibodies for the diagnosis, prevention of treatment of whooping cough.		

XX Sequence 11883 BP; 2196 A; 3648 C; 4282 G; 1754 T; 3 other;
 Alignment Scores:
 Pred. No.: 1.31 Length: 11883
 Score: 143.50 Matches: 112
 Percent Similarity: 39.43% Conservative: 80
 Best Local Similarity: 23.00% Mismatches: 205
 Query Match: 5.95% Indels: 91
 DB: 21 Gaps: 21
 US-09-825-414-7 (1-486) x AA10263 (1-11883)
 QY 33 SerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHis 52
 DB 1492 GCGGGGAGAGCGCGTCGCTCCGGAACGCGGCGCGCCCTCAAGCGGACAG 1551
 QY 53 ValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaValAlaAsp 72
 DB 1552 CTGTGGCGAGCGCGGCTC-----GACGTGGATGGCAAGCAGGCGCTCGCG 1599
 QY 73 ValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAla----- 86
 DB 1600 CTGGGGTGGCCAGCAGCATGCGCTG---TCGGTCCTGTCGCGCGCGCCCTCAAGCGC 1656
 QY 87 -----SerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
 DB 1657 GGCAAGCTGTCGGCGAGCGGCGGCTGACGTGACGAGCAGCAGCGCTCACGCTGGGT 1716
 QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
 DB 1717 TCGGTTCGAGCGAGCGT-----GCGCTGTGCGTAAGCGCTGGCGAAACCTG 1764
 QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMet---GlnProAlaIleAsn 139
 DB 1765 GCGGCGAAGCATGTGCTCCACGTGCCCACTTGAAGTCCGTGGCGAGCGGAGGTGCGG 1824
 QY 140 LysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeu 159
 DB 1825 CTGATGAC-----GCTTCGAGCGACGCGCGCATGACCGGTGTCCGACGAGCGCTG 1878
 QY 160 Ser-----GlyAlaMetAspGlnValGly----- 167
 DB 1879 GCGGCGCGCACCTCGACGTCCAGGCGCCATCGCGGTACAGGCTGAGAGCGGCTCAC 1938
 QY 168 -----ThrLysMetMetAspArgAlaThrArgLysAspLeu 178
 DB 1939 GTGGCGCAACGACACGACCGCATTCGCGTCCGCGGCGCGCGCCAGGTGATCTG 1998
 QY 179 HisTyrLeuSerAlaSer-ProAspArgLeuHisAspAlaMetAlaAlaSerValLysAr 198
 DB 1999 CACGACTGAGCGAGCGCGCGCGGATATCTCCGCGAGGGGCGCTCATATACGCG 2058
 QY 198 gHisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSe 218
 DB 2059 CGTGGCGCGCATGATGATGTAAGTCTCCGCGCACGCGCGCTGTGTCGATTCGATAC 2118
 QY 218 fAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln 238
 DB 2119 ATGAGCGCGCTGCTGCTGCTGCGCTCCAGCGAGCGGAGCGGTGTCGCGCAAGCATAG 2178
 QY 238 nGlyAlaValAspLeuGlyValSerMetAlaGlyLeuAlaAlaAlaAsnAlaGlyPheGln 258
 DB 2179 CCGACCGCTGGC-----GCCGTACCGTCAAGCGGCGCGCTCAACCTGCGC----- 2229
 QY 258 yAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGly---GlyAlaLeuValLe 277
 DB 2230 -----GATGTCCATCGATGCGGAGCGAGTCCGCGCACCGAGCGCGCGCTTGA 2276
 QY 277 uGlyLeuLysAspLysGluProLysAlaGlnLeuSer---GluGluAsnAspTyrLeuGln 296
 DB 2277 GACGCTGCGAGACGTGCGGCTGCGCGCACCTTGCCTGCGACGCGGCGGACCGCTTGA 2336

QY 296 uAla-----TyrLysAlaIleLysSerAlaSerTyrSerGlyAla---AlaLeuAsnAl 313
 DB 2337 GCGCGGGTCTCGAATATCGCGCGGTCCATGACCGCTGAACGCGCGCATGCTCGACT 2396
 QY 313 aGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValAr 333
 DB 2397 GGATGCGCGCGAGCGGGGGGCAATGCGGGTTCACAGCGGCGAGGCTGCTGGG 2456
 QY 333 gSerLeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAl 353
 DB 2457 CAGTCTCGGCGCGAGGCGGAGCTGACGCTATCG-----GCCGCGCGCGCGC 2504
 QY 353 aGlyValGlyLysLeuGlnMetAlaThrLysAsnIleThr----- 367
 DB 2505 GACCGTGGCGGATGAGTTCGTGACACATCTCCGTGACGCGGCGGAGCGGCTGTGTC 2564
 QY 368 -----AspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeu----- 381
 DB 2565 GGTTCAGACGCTCAACAGCGGCTCCAGGCTCGCATTTGCGCGCACGCGCGCTGATGT 2624
 QY 382 -----AlaGlySerAlaAlaValPheAlaGlyTyrThrAla----- 394
 DB 2628 AGCAAGGTTCGCGCAGAGCGGTATCGGCTCGAAGGCTGGCGGCTCGAGCGGA 2684
 QY 395 -AlaLeuThrThrAspProAlaValLysLysAla-----GluSerPheIleGlnAspTh 412
 DB 2688 CTCCCTCGGTTCGAGCGCGCGCATGACGCTGCGGCGCGATCGGTACAGGTGATCA 2744
 QY 412 rValLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThr-----Va 429
 DB 2745 AGCCCGCATGTCCTGCGCATTTGCTGGGGGCGGAAGCGCGCGCACGCTGCGCGCTG 2804
 QY 429 LysLeuAlaLysThrValLysAspMetGlyGlyGlnAlaIleThr-----Hi 445
 DB 2805 GGAGCGCGCGGTGTGATGATGACGTGCGCGCGATCCACGATCGGGGAGTCTGCTGCA 2864
 QY 445 sThrGlyAlaSerLeuArg 451
 DB 2865 CCGCAATCGCGAGTTCGG 2883
 RESULT 10
 ID AAZ20203 standard; DNA; 1801 BP.
 AC AAZ20203;
 DT 17-JAN-2000 (first entry)
 DE Mycobacterium tuberculosis antigen fusion protein Mtb61f DNA.
 KW Tuberculosis; antigen; fusion protein; Mtb61f; TbH9; DPV; MTI;
 KW diagnosis; therapy; vaccine; immunogen; ss.
 OS Mycobacterium tuberculosis.
 PN W09051748-A2.
 PD 14-OCT-1999.
 PF 07-APR-1999; 99WO-US07717.
 PR 07-APR-1998; 98US-0056556.
 PR 30-DEC-1998; 98US-0223040.
 PA (CORI-) CORIXA CORP.
 PI Skelky YAM, Alderson M, Campos-Neto A;
 DR WPI; 1999-601610/51.
 DR P-PSDB; AAY32068.
 PT New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis -

XX Example: Fig 10A-B; 83pp; English.

CC This DNA sequence includes a coding region for a recombinant
CC Mycobacterium tuberculosis tri-antigen fusion protein (see AAY32068),
CC termed Mtbf1, composed of the antigens Tbh9, Dpv and MT1. The
CC DNA is useful for the recombinant production of the fusion protein.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein. Coding sequences for Tbh9, Dpv and MT1 were then ligated
CC to encode Mtbf1. The invention provides fusion proteins (see
CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as
CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
CC in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components.

XX Sequence 1801 BP; 288 A; 558 C; 642 G; 313 T; 0 other;

Alignment Scores:

Pred. No.:	0.157	Length:	1801
Score:	143.00	Matches:	117
Percent Similarity:	34.45%	Conservative:	58
Best Local Similarity:	23.03%	Mismatches:	165
Query Match:	5.93%	Indels:	168
DB:	20	Gaps:	25

US-09-825-414-7 (1-486) x AAY20203 (1-1801)

QY 20 ThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln 39
DB 262 ACCGGCGGGCAGCGAGTGAACCGCGCCAGAGTCCGGTT----- 303
QY 40 ArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaIleHisLysLeu 59
DB 304 -----GCTGGCGGGCTTACGAGCGCGCTATGGCTGACG-----GTG 342
QY 60 ProProAlaAspSerAlaSpGlyGlnAla----- 69
DB 343 CCCCAGCGCGGTATGCCCGAGAACCTGCTGAACCTATGTTCTGATACGACCAACCTC 402
QY 70 -----AlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGlu 83
DB 403 TTGGGCAAAACACCCCGCGATCGCGTCAACGAGCGCGAATACGCGAGATGGGCC 462
QY 84 ThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPhe 103
DB 463 CAAAGACGCGCGCGATGTTGGCTAAGCGCGGCGAGCGGCGAGCGGCGAGCGGCTTG 522
QY 104 AlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu----- 120
DB 523 CTGCCCTTGAGAGAGCGCGGAGATG-----ACCAAGCGGGGTGGCTCTCGAGCAG 576
QY 121 -----ArgAlaThrProPheAlaMetAlaSer---LeuLeuGlnTyrMet 134
DB 577 GCCGCGCGGTGAGAGAGCGCTCCGACCGCGCGCGAACAAGTGAATGAACAATGTG 636
QY 135 GlnProAlaIleAsnLysLysAspTrpLeuProAlaProLeuLysProLeuThrProLeu 154
DB 637 CCCCAGGCGCTGCACAG-----CTGGCCACGCCACAGAGGCGACACAGCGCT--- 684
QY 155 IleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly-----Thr 168
DB 685 ---TCTTCAAGCTGGGTGGCTGTGGAAGAGCGGTCTCCCGCATGGTCCCGCATCAGC 741
QY 169 LysMetLysAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeu 188
DB 742 AACATGTGTGATGGCCAAACACATGTGCG---ATGACCAACATCGGGTGTGTGATG 798
QY 189 HisAspAlaMetAlaIleAspValLysArgHisSerProSerLeuAlaGlnValLeu 208
DB 189 HisAspAlaMetAlaIleAspValLysArgHisSerProSerLeuAlaGlnValLeu 208

DB 799 ACCAACCTTGTAGCTCATGTTGAAGGCTTGTCTCCGGCGCGCCGCCAG----- 852
QY 209 AspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla 228
DB 853 -----GCCGTGCAAAACC---GCCGCCAAAGCGGCTCCG----- 885
QY 229 ProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyAlaSerMetAla 248
DB 886 ---GCCATGAGCTCGCTGGCGACCTCGCGGTCTTCGGGTGGC----- 930
QY 249 GlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAsp 268
DB 931 GGTGGGTGGCGCCCAACTTGGGT-----CGGGCGGCTCGGTGGTTCG----- 975
QY 269 HisGlnArgGlyAlaLeuValLeuGlyLeuLysAspLysLurProLysAlaGlnLeu 288
DB 976 -----TTG 978
QY 289 SerGluGluAsnAspTrpLeuGlnAlaTyrLysAlaIleLysSerAlaSer----- 305
DB 979 TCGGTGCCGAGCGCTGGCGCGCGCGCCACACAGCATCACCAGCGCGCGCGCGCTG 1038
QY 306 -----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB 1039 CCGCTGACCAAGCTGTACACCGCGCGGAAAGAGCGCCGCGCATGCTGGCGGCTG 1098
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
DB 1099 CCGGTGGC----- 1107
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal----- 355
DB 1108 ---CAAGTGGCGCCAGCGCGCGGTGGTGGCTCAGTGTGTGCTGGGTTCGCGCG 1161
QY 356 -----GlyLysLeu----- 358
DB 1162 CGACCTATGTATGCCGATTCCTCCGACCGCGCAAGCTGTATCCCGTGAAGCGGCTC 1221
QY 359 -----GlnGluMetAlaThrLysAsnIleThrAspPro--- 369
DB 1222 ATTACACCACTGCATTAACGGGCGAGTACGTACGCTCAACCGCAGCATCCGCGG 1281
QY 370 -----AlaThrLysAlaAlaValSerGlnLeuThrAspLeuAlaGly 383
DB 1282 GCTGCCGCAAGTTCAACGCGTCAACGCGGCGCGACGCTCATATTGGCAATTCCGCGC 1341
QY 384 SerAlaIleValPheAlaGlyTrpThrAlaAlaLeuThrThrAspProAlaValLys 403
DB 1342 GCACCGCCACCTCAGCGCGCTGCCATGCCGCCAATTG-----CAAGCTGTGCG 1392
QY 404 LysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyr 423
DB 1393 GGGGCGGCAAGTACATC---GGCTTGTGAGTCTGCTGGCTCGCTCGCAACATCAT 1449
QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlnAlaIle 443
DB 1450 GAG-----CTCATGACGATTAATTACAGATTGGGAGCGTCGAC 1488
QY 444 ThrHisThrGlyAlaSerLeuArg 451
DB 1489 GCTCAT---GGCGCATGATCCG 1509
RESULT 11
AAK51943
ID AAK51943 standard; cDNA; 7545 BP.
XX
XX AAK51943;
AC AAK51943;
XX
XX
DT 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 488.
DE Human polynucleotide; cell proliferation; cell differentiation; gene therapy;
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; hematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0634936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QH, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejherman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR P-PSDB; AAM78810.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
XX Claim 1; Page 1755-1761; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAW80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, hematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibitor activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 7545 BP; 1856 A; 2006 C; 1941 G; 1742 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.912 Length: 7545
Score: 142.50 Matches: 125
Percent Similarity: 37.55% Conservative: 74
Best Local Similarity: 23.58% Mismatches: 250
Query Match: 5.91% Indels: 82
DB: 22 Gaps: 22
US-09-825-414-7 (1-486) x AAK51943 (1-7545)
QY 10 PropovalThralaThraspSerPhearThralaSerAspAlaSerLeuAlaSerSer 29
DB 885 CCATGATGAGCATCAACCGGTGCATCCGGACATGACAGAG-GCCTCGGTGGCCGCCGC 943
QY 30 SerValArgSerValSerSerAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 49
DB 944 AGC--CAGAGCGCTGGCCGAGGAGCAGCATCTCTGTGAGAGCCCTGCAGAGAGAGCTG 1000
QY 50 ThrAspHisValPheAlaAlaHisLysLeu-----ProProAlaAspSerLysAspGly 67
DB 1001 ACTTCGGTGTCTCAGAGAAATCGGACACCTTATTCATCCATCGACACAGCGGCTCGGGGA 1060

QY 68 GlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
DB 1061 GAAGCAGCTCAGCTGGACAT-----AAGTGACACAACTGGCAAG----- 1102
QY 88 ArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGlu 107
DB 1103 -----TATTTTGGAG-----CCCTGATCTTAGCCGAGTGGTGGCCCTCAAG 1147
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuAlaGlyAlaThrProPheAlaMet 127
DB 1148 ATTCCTGATCATCAGCAGCAGATGAGC-----GTCTGGACCAAGCAAGACTCTCGCA 1201
QY 128 AlaSerLeuLeuGlnIleThyMetGlnProAlaIleAsnGlyAspThrLeuProAlaPro 147
DB 1202 CAGTCTGCCTTCAG--ATGTTGATGACGCAAGAAAGAGTGGGGAAGCCCAAGCA 1258
QY 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyAlaMetAspGln 165
DB 1259 CAACACACCCATGAGCCCATCAGAGGCGCCGAGTTGATGAAGAGCCGTGGATGAC 1318
QY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisLysLeuSerAlaSerPro 185
DB 1319 ATC---ATGTCAGCCTGCAACGACCTGCCAGTGAAGTGGGCTGTGGGCGCATGTG 1375
QY 186 AspArgLeuHisAspAlaMetAlaIleSerValLysArgHisSerProSerLeuAlaArg 205
DB 1376 GAGCCCATTTGCAAGAGCCATGAGCAAGCTGATGAAGCAGCTCTCCAGAACCAAGGGA 1435
QY 206 GlnValLeuAspThrGlyValAlaValGlnThrLysSerAlaArgAsnAlaVal----- 223
DB 1436 ACATTTGTCGACTATACAGACGACTGTGTAATCTCCAAAGACTTGGCTGACAGCT 1495
QY 224 -----ArgThrValLeuAlaProAlaLeuAlaIleSerAlaArgProAlaValGln 238
DB 1496 CAGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1555
QY 239 GlnAlaValAspLeuGly---ValSerMetAlaGlyGlyLeuAlaAla----- 253
DB 1556 ATGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1615
QY 254 -----AsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis----- 269
DB 1616 CCAGAGGAGATGCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT 1672
QY 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
DB 1673 ATCTTCCTGTGTCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1732
QY 286 AlaGlnLeuSerGlu---GlnAsnAspThrLeuGlnAlaLysAlaIleLysSerAla 304
DB 1733 AGGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1792
QY 305 SerTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
DB 1793 CTCAGAGCCGCGGACAAAGAAAGAACCCAGGATTCATTAACAGCCGCTGTGCTGGG 1852
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValAlaArgSerLeuValSerAlaSer 339
DB 1853 ATCATTTGCCGAGCTGGACACCACTTATGTTTGGACAAGCGGGAGAGCTGAATCAGAG 1912
QY 340 SerLeu-----ThrGlnAsnGlyLeuAlaIleAlaGlyLysIlePheAla 353
DB 1913 AACAGTGAAGACTTGGCAGACACAGAGGAGAACATTTCAAGACGCGCAAGGCTTGGTA 1972
QY 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
DB 1973 GAAGACAGAAACTACTTGTGTGTCAGAGACTGGCTCCTCCTCAAGACTGGCCCAAGCG 2032
QY 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValAlaValAlaGlyLysThrThr 393
DB 2033 GCCAGTCTCTCAGCAGCCACCATCAACCAAGCTCCGAGAAAGTGTCAAGCTGGGGCAGCC 2092

QY 394 AAlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
DB 2093 AACCTGGGGTCCGACGACCCGAGACCCAGTGGTTTGATCATGCCATCAAAAGTGTG 2152
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrGlyValAla 425
DB 2153 GCCAAGGCCCTTCTGATCTCATCCAGTCTCCAAAGGAGCTCCAGCAAGCAGTGGAC 2212
QY 426 AspGlnThrVal-----Lys 430
DB 2213 GACCTTCATCTACCAAGTCAAGGGGCTGCCAAGGTATGTGACCAATGCACCTGG 2272
QY 431 LeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHisThrGlyAlaSerLeu 450
DB 2273 CTCCTCAAGACTGTAAAGCAGTGAAGATGAGCACCAGGGGCGCAGGCGG---CTT 2329
QY 451 ArgAsnThrValAsnAsnLeuArgGlnArgProIle-----ArgGlnAlaAspIleGlu 468
DB 2330 GAGGCCACAAATTGAATGCATTAAGCAGAGCTTACGGTGTCCAGTCAAAAGACGTACCT 2389
QY 469 GluGlyGlyThrAlaAlaSerProSerGlu 478
DB 2390 GAA-----AAGCATCATCACCCTGAAGA 2413
RESULT 12
AA158449
ID AA158449 standard; cDNA; 7615 BP.
AC AA158449;
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 652.
DE Human polynucleotide SEQ ID NO 652.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KM peripheral nervous system; neuropathy; central nervous system; CNS;
KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
KM Leukaemia; ss.
XX Homo sapiens.
OS Homo sapiens.
PN WO200153312-A1.
XX 26-JUL-2001.
PD 26-DEC-2000; 2000WO-US34263.
PF 26-DEC-2000; 2000WO-US34263.
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HISE-) HISEQ INC.
PA
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM39293.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 652; 10078bp; English.
PS The invention relates to human nucleic acids (AA157798-AA161369) and
CC

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 7615 BP; 1874 A; 2026 C; 1954 G; 1754 T; 7 other:
Alignment Scores:
Pred. No.: 0.922 Length: 7615
Score: 142.50 Matches: 125
Percent Similarity: 37.558 Conservative: 74
Best Local Similarity: 23.588 Mismatches: 250
Query Match: 5.91% Indels: 82
DB: Gaps: 22
US-09-825-414-7 (1-486) x AA158449 (1-7615)
QY 10 PROFOVALTHRALATHRASPSERPHRARGTHRALASERASPSERLEUALASER 29
DB 955 CCATCGATGACATCAACCGGTGATCCGGGACATCGACAG---GCCCTGGCGCCGCCCTC 1013
QY 30 SerValArgSerValSerSerAspGlnArgGluIleAsnAlaIleLeuAlaSer 49
DB 1014 AGC---CAGAGCCTGGCCACGAGGACGACACATCTGTGGAGGCTCGAGGACGCTG 1070
QY 50 ThrAspIleValPheAlaIleHisLysLeu-----ProProAlaAspSerAlaAspGly 67
DB 1071 ACTGTGGTGTGCAGGAATTCGGACACCTTATTCGATCCATCGCCACACGCGTGGGGA 1130
QY 68 GlnAlaIleValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
DB 1131 GAAGCAGCTCAGCTGGGACAT-----AAGGTACACAACTGGCAAGC----- 1172
QY 88 ArgLeuHisPheGluGluGlyGluThrProAlaThrIleAlaAspThrPheAlaSer 107
DB 1173 -----TATTTGAG-----CCCTTGATCTAGCCGCACTGGTGGCTCCAAAG 1217
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
DB 1218 ATTCTTGATCATCAGACGACGATGAGC-----GTCTGGACCAACGACACTCTCGCA 1271
QY 128 AlaSerLeuGlnIleThrMetGlnProAlaIleAsnLysGlyAspThrLeuProAlaPro 147
DB 1272 GAGTGTGCTTGGCAG---ATGTTGTATGCAGCCAAAGAGGTGGCGAAACCCCAAGGCA 1328
QY 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerLysAlaMetArg 165
DB 1329 CAACACACCCATGACGCGCATCACAGAGCGCCGCACTGATGAAGAACCCGTGGATGAC 1388
QY 166 ValGlyThrLysMetLeuAspArgAlaThrGlyAspLeuHisIleThrLeuSerAlaSerPro 185
DB 1389 ATC---ATGTGACGCTGAACGAAAGTGGCAGTGAAGTGGGTGGGGGCAATGGTG 1445
QY 186 AspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArg 205
DB 1446 GAGCGCATTTGACAGAACCATGAGCAGCTGGAGGAAGGCACTCTCCAGAACCAAGGGA 1505
QY 206 GlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal----- 223
DB 1506 ACATTGTGTGACATTCAGACGACTGTGTAAATCTCCAAAGCAATTCGCGGTGACAGCT 1565
QY 224 -----ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaGln 238

```
Db 1566 CAGGAATGATGACTTAAGTCGGTTACTAACCAGGAGGATTGGAGAGACTGCTTCACAA 1625
QY 239 GtYlaValaLAspLeuGly----ValSerMetAlaGlyGlyLeuAlaAla----- 253
Db 1626 ATGACCACTGACTATGACCTGGCTTCCAGGCGCAAGATGGCCAGGCCACGGCGGAA 1685
QY 254 -----AsnAlaGlyPheGlyAsnArgLeuSerValGlnSerArgAspHis----- 269
Db 1686 CCAGAGAGATCGGATTCACATTGCG---ACTGCTGTGCAGAGACTTGGCCACGGCGGT 1742
QY 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGlnProLys 285
Db 1743 ATCTTCTGTGTGCAGAAAGCAGGGCGCTCCAGCTGTGCCCCACAGACAGCTACACCAAG 1802
QY 286 AlaGlnLeuSerGlu---GluAsnAspTrpLeuGlnAlaIleValAlaIleLysSerAla 304
Db 1803 AGGGAGCTGATCGAATGCGCCCGCTGCGTACAGGAAAGTCTCTTGGTGTCTGTGGCT 1862
QY 305 SerTrpSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
Db 1863 CTCACGGCGGGGAACAAGAACCCAGGCAATGCATTACAGCCGCCACGGCTGTGTGG 1922
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSer 339
Db 1923 ATCATTTGCCGACCTGGACACACCATATATGTTCGACACAGCGGGGAGCGCTGATGCAGAG 1982
QY 340 SerLeu-----ThrGlnAsnGlyLeuAlaLeuAlaGlyGlyLysPheAla 353
Db 1993 AACAGTGAAGCTTCGACGACACAGGAGAGAACATTCCTCAAGACGGCCAAAGCCTTGTA 2042
QY 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
Db 2043 GAAGACACGAACACTGTTGTGTGCAGAGCTGCTGCTACTCTCTCAACAACTGGCCAGGCG 2102
QY 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThr 393
Db 2103 GCCCAGTCTCTAGCAGCCACCATCACCAGCTCGCAGAAAGTGTCAAGCTGGGGCAGCC 2162
QY 394 AlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
Db 2163 AGCGTGGCTCCGACGACCCGAGACCCAGGTGTTTGTATCAATAGCCATCAAAATATGTG 2222
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrGlyTyrValAla 425
Db 2223 GCCAAGGCCCTTTCATCTCATCTACGTCTACCAAGGAGGCTGCCAGCAAGCCAGTGCAGC 2282
QY 426 AspGlnThrVal-----Lys 430
Db 2283 GACCTTCATGATTCACCACTCAAGGGGCTGCCAAGGTAGTGTGACCAATGTCACTCG 2342
QY 431 LeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHisThrGlyLysSerLeu 450
Db 2343 CTCCTCAAGACTGTAAAGCAGTGGAGATGAGGCCACCGGGGACACAGGGCG---CTT 2399
QY 451 ArgAsnThrValAsnAsnLeuArgGlnArgProAla-----ArgGlnAlaAspIleGlu 468
Db 2400 GAGGCCCAATGTGATGTAAAGCAGAGCTTACGTTCTTCAGTCAAAAGACGTACCT 2459
QY 469 GluGlyGlyThrAlaAlaSerProSerGlu 478
Db 2460 GAA-----AAGACATCATCTACCTGAAGAA 2483

RESULT 13
AA158468
ID AA158468 standard; cDNA; 7657 BP.
AC AA158468;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 671.
XX
```

```
KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia; ss.
XX
XX Homd sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-052317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QA, Zhou F, Goodrich R, Drmanac RT;
XX WPI: 2001-442253/47.
XX P-PSDB: AAM39312.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Claim 1: SEQ ID NO 671; 10078bp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localized neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX SQ Sequence 7657 BP; 1889 A; 2036 C; 1971 G; 1761 T; 0 other;

Alignment Scores:
Pred. No.: 0.928 Length: 7657
Score: 142.50 Matches: 125
Percent Similarity: 37.55% Conservative: 74
Best Local Similarity: 23.58% Mismatches: 250
Query Match: 5.91% Indels: 82
DB: 22 Gaps: 22

US-09-825-414-7 (1-486) x AA158468 (1-7657)
QY 10 PropValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
Db 997 CCATGATGATCATCAACCGGTGATCCGACATGACAG-GCCTTCCTGGCGCCGCTC 1055
QY 30 SerValArgSerValSerSerAspGlnGlnArgGlnIleAsnAlaIleAlaAspTrpLeu 49
|||
||| ::::::::::::::::::::
||| ::::::::::::::::::::
```

```

Db 1056 AGC---CAGAGCCTGGCCAGAGGACGACATCTCTGTGGAGCCCTGCAGGACGACGCTG 1112
Qy 50 ThrAspHisValPheAlaIleHisLysLeu-----ProProAlaAspSerAlaAspGly 67
Db 1113 ACTTGGCTGGTGCAGAGAAATCGGACACCTTATTCGATCCATCGCACACGGCTGGGGGA 1172
Qy 68 GlnAlaIleValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
Db 1173 GAAGCAGCTCAGCTGGGACAT-----AAGTGACACCAACTGGCAAGC----- 1214
Qy 88 ArgLeuHisPheGluGluGlyLutThrProAlaThrIleAlaAspThrPheAlaLysAlaGlu 107
Db 1215 -----TATTTTGAAG-----CCCTTGATCTTAAGCCGACGTGGTGGCTGGCTCAAG 1259
Qy 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
Db 1260 ATTCTTGATCATCAGCAGACGATGACG-----GTCTGGACCACGACCAAGACTCTCGCA 1313
Qy 128 AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuProAlaPro 147
Db 1314 GAGTCTGCCTTGCAG---ATGTTGATATGCAGCCAAAGAGTGGCGGAACCCCAAGGCA 1370
Qy 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyAlaMetAspGln 165
Db 1371 CAACACACCCATGACGCCATCACAGAGCGCCAGCTTGATGAGAAGAACCGCTGATGAC 1430
Qy 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
Db 1431 ATC---ATGTCAGCCTGAACAGACAGACTGCCAGTGAAGTGGCTGGTGGGCGCATGTG 1487
Qy 186 AspArgLeuHisAspAlaMetAlaSerValLysArgHisSerProSerLeuAlaArg 205
Db 1488 GACGCCATTCGACAGACCCATGAGCAAGCTGATGAAAGCACTCTCAACACCAAGGGA 1547
Qy 206 GlnValLeuAspTrpGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal----- 223
Db 1548 ACATTTGTGACTATCAGACGACTGTGTAAATACTCCAAAGCACTGCGGTGACAGCT 1607
Qy 224 -----ArgThrValLeuAlaProAlaLeuAlaSerArgTrpArgValAlaGln 238
Db 1608 CAGGAATGATGACTAAAGTGGTACTAAACCGGAGAGTGGGAGACTGGCTTCACAA 1667
Qy 239 GlyAlaValAspLeuGly---ValSerMetAlaGlyLysLeuAlaIle----- 253
Db 1668 ATGACACAGTACTATGGACACTGGCTTCCAGAGCCACATGGCAGCAGCGGGA 1727
Qy 254 -----AsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis----- 269
Db 1728 CCAGAGAGATCGGATTCAGATTCCGATTCGC---ACTCGTGCAGAGACTGGCGCAGGCTGT 1784
Qy 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
Db 1785 ATCTCTCTGGTGCAGAAAGGAGGAGGCGCTCCAGAGTGGCCCAACACAGCTAACACCAAG 1844
Qy 286 AlaGlnLeuSerGlu---GluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAla 304
Db 1845 AAGGAGCTATGAAATGGCCCGCTGCCCTCAGCGAAAGAGTCTCTGTGGCTCTGGCT 1904
Qy 305 SerTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
Db 1905 CTCACGCGCGGGAACAAAGAACCCAGCGATCATTAACCGCGCACCCCTGTGGTGG 1964
Qy 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSer 339
Db 1965 ATCATTTGCCGACCTGGACACCACTATTATGTTTGCMAACAGCGGGAGCGGTGATGACGAG 2024
Qy 340 SerLeu-----ThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAla 353
Db 2025 AACACGACACCTTCGACACACCAAGGAGAACATTTCTAAGACGGCCCAAGGCTTGGTA 2084
Qy 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
Db 2085 GAAGACACGAAACTACTTGTGTACAGAGCTGGCTCCACTCTGACAAAGCTGGCGCCAGGGG 2144

```

```

Qy 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThr 393
Db 2145 GCCCACTCTCTCAGACGACCCACATCACCCAGCTCGCGAAGTGTCAAGCTGGGGCAGCC 2204
Qy 394 AlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
Db 2205 AGCCTGCGTCCGACGACCCCGAGACCCAGTGTGTTTGTATCAATGCTCAAAAGATGTG 2264
Qy 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAla 425
Db 2265 GCCAAGGCCCTTCTGATCTCATCACTGCTACCAAGGAGGTGCCAGAACGACGTGAC 2324
Qy 426 AspGlnThrVal-----Lys 430
Db 2325 GACCTTCATGATGACAGCTCAAGGGGCTGCCAAGTATGTGATGCAATGCTCACTCG 2384
Qy 431 LeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeu 450
Db 2385 CTCCTCAAGACGTAAAGCAGTGGAGATGAGCCACCGGGGCGACAGGGCG---CTT 2441
Qy 451 ArgAsnThrValAsnAsnLeuArgGlnArgProAla-----ArgGluAlaAspIleGlu 468
Db 2442 GAGGCCACATGTAATGATGATAAGCAGAGCTTACGCTGTTCACGTCAAAAGACGTACT 2501
Qy 469 GlnGlyGlyThrAlaAlaSerProSerGlu 478
Db 2502 GAA-----AAGACATCATCACTCGTAAGAA 2525

RESULT 14
ID127385
ID127385 strand: DNA; 2466 BP.
AC ABL27385:
XX
XX
XX 26-MAR-2002 (first entry)
DX
XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 33628.
KW Drosophila: developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
PN WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PR
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX Venter JC, Adams M, Li PMD, Myers EW;
PI WPI; 2001-656860/75.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Claim 1; SEQ ID NO 33628; 21pp + Sequence Listing; English.
CC
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

```


QY 272 Gly-----GlyAlaLeuValLeuGlyLeuLysAspLys-----GluProLys 285
 ||||| |||::: :::||||
Db 35049 GCTGACCGCGGAGCGCTGCGCGCAAGGTGTGCGCCCAAGGTGCAGCGCGGCCAACCACT 35108
QY 286 Ala-GlnLeuSerGlnGluAsnAspTrpLeuGluAlaTyrLysAlaLeuLysSerAla 305
 :::||||: ::: |||:::||||
Db 35109 CGAGGAGCTGACCGCGCACTGGAC---CTGACGCGGTTCATGCTACTGCTCGCCCTC 35165
QY 305 r-----TyrSerGlyAla-----Al 310
 | |||::: |||
Db 35166 CTGATCTTCATGGCGCGCGCGCGCGCGGTACGCGCGCGCAACGCTACCTCGACGG 35225
QY 310 aLeuAsnAlaGlyLysArgMetAlaGlyLeuPro---LeuAspMetAla----- 325
 ||| :::|||| ||| :::||||
Db 35226 CCTGATGGCCCGCGCGCGCGCGCGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCG 35285
QY 325 ----- 325
Db 35286 GGACGAGCTCACCGGATGCGCGCGACACCATCGACGACTCACCTGGCCCGGATGAGCCG 35345
QY 326 -ThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerLeuThrGlnAsnG1 345
 ::: ||| |||::: :::||||
Db 35346 GCGGGAAGCGCGCGCGCGCGCGCTGCGCGCTGCGCGCGCGCGCATGAGCTGTTCGA 35405
QY 345 yLeuAlaLeuAlaGlyGly----- 351
 ||||| |||
Db 35406 CGCGCGCTCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 35465
QY 352 -----PheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAs 368
 ||||| ||| ::: :::||||
Db 35466 GCGGCG 35522
QY 368 pProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValAlp 388
 :::||||| ::: |||
Db 35523 CGCGGCG 35582
QY 388 eAlaGlyTrpThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPh 408
 ||||| ||| ::: |||:::||||
Db 35583 GGCGGCGCTCACCGTGGCC-----GNAACAGAAAGCGCT 35615
QY 408 eIleGlnAspThrValLysSerThrAlaSerSerThrGlyTyrValAlaAspGlnTh 428
 ::: ||| |||::: |||:::||||
Db 35616 GCTGCTGACCTCTGCTCGCGGTCAAGTGCAGTGCCTGCGGCAACGCGACAGCTCCGG 35675
QY 428 rValLysLeuAlaLysThrValLysAspMetGlyGluAlaIleThrHisThrGlyAl 448
 |||::: ||||| ||| :::||||
Db 35676 CGTCG 35729
QY 448 aSerLeuArgAsnThrValAsnAsnLeu--ArgGlnArgProAlaArgGluAlaAspIle 467
 ||||| ::: ||| |||:::||||
Db 35730 GGAAGTGGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35789
QY 468 -----GluGluGlyGlyThrAlaAlaSerProSer 477
 ||| ||||| :::||||
Db 35790 CGACCATCGAAACCGCGCTGGCACTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 35828

Search completed: January 31, 2003, 05:29:09
Job time : 3263.75 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 03:21:14 ; Search time 1572.38 Seconds

(without alignments)
5005.784 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412

Sequence: 1 MHNRVQGPVATDSEFR.....IEEGTAASPEIFRRPMS 486

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+g2n.model -DEV=xip
-O=/cgn2.1/USPTO.spool/US09825414/tunat_28012003.161742.26093/app-query.fasta.1.1294
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptio -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09825414_EGCG_1.1.2221.6runat_28012003.161742.26093 -MCPu=6 -ICPU=3
-NO_XLUPX -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	126.5	5.2	1138	17	CNS060E0
2	126	5.2	1753	17	BC018800
3	125.5	5.2	4001	11	AK014542
C 4	119	4.9	2191	17	A0012154
C 5	115.5	4.8	883	12	BC369810
C 6	114	4.7	1792	11	AY103948
7	113.5	4.7	4088	11	BC021514
8	112.5	4.7	2012	11	AY107104
C 9	111.5	4.6	975	9	AL545051
C 10	111	4.6	948	13	BM044579
C 11	110.5	4.6	711	13	BM588321
C 12	110.5	4.6	888	17	CNS076CP
C 13	110	4.6	860	14	BQ424977
14	109.5	4.5	670	14	BQ027919
15	109.5	4.5	2347	11	BC020083
16	109	4.5	671	13	BT725871
17	109	4.5	1295	12	BF973453
18	107	4.4	3264	11	AK004694
19	106.5	4.4	1418	11	AK019388
20	106	4.4	2372	13	BM042670
21	105.5	4.4	1038	17	CNS029PG
22	105	4.4	989	14	BQ065543
23	104.5	4.3	1059	13	BM57363
C 24	104.5	4.3	1108	13	BM548553
25	104.5	4.3	1150	17	CNS079ZS
C 26	104.5	4.3	1603	11	AY103757
27	104.5	4.3	2802	11	BC023619
28	104.5	4.3	3441	11	BC030898
29	104	4.3	660	13	BT476386
C 30	104	4.3	665	13	BQ841850
C 31	104	4.3	915	14	BQ952432
32	104	4.3	1159	17	CNS076BD
C 33	104	4.3	1923	11	AY106615
C 34	103.5	4.3	718	13	BT523897
C 35	103.5	4.3	759	9	AL497335
C 36	103.5	4.3	1033	14	BQ685858
37	103	4.3	774	13	BT556804
38	103	4.3	910	14	BQ688626
C 39	103	4.3	1002	17	AG113524
C 40	103	4.3	1031	12	BG482132
41	103	4.3	1979	11	AY104979
C 42	103	4.3	4272	11	BC016144
43	102.5	4.2	732	14	BQ804418
44	102.5	4.2	746	17	A2624776
45	102.5	4.2	950	9	AL541558

ALIGNMENTS

RESULT 1
CNS060E0/c
LOCUS
DEFINITION T3 end of clone AW0AA005C02 of library AW0AA from strain CLIB 89 of
ACCESSION ALA10542
VERSION ALA10542.1 GI:12179004
KEYWORDS
SOURCE
ORGANISM
Yarrowia lipolytica.
Yarrowia lipolytica.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE
1 (bases 1 to 1138)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Boitlot,Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,

[illegible][illegible]

	VERSION	AKO15442.1 GI:12852462
KEYWORDS	HNC; CAP trapper.	
SOURCE	Mus musculus (Strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone_11b:RIKEN full-length enriched mouse cDNA library clone:4631434016.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia, Eutheria, Rodentia, Sciurognathi, Muridae, Murinae, Mus.	
AUTHORS	Carninci,P. and Hayashizaki,Y.	
TITLE	High-efficiency full-length cDNA cloning	
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	99279253	
PUBMED	10349636	
REFERENCE	2	
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE	20499374	
PUBMED	11042159	
REFERENCE	3	
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,T., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi.N., Ishii.Y., Nakamura.S., Hazema.M., Nishino.T., Harada.A., Yamamoto.R., Matsumoto.H., Sakaguchi.S., Ikegami.T., Kashtwagi.K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawaj,J., Okazaki,Y., Muramatsu,M., Inoue.Y., Kira,A. and Hayashizaki,Y.	
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multipipillary sequencer	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PUBMED	11076861	
REFERENCE	4	
AUTHORS	Kawai,T., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arikawa,T., Hara,A., Fukunishi,Y., Kono,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamana,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadoie,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiya,H., Knehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush.J., Schriml,L.M., Strubill,F., Suzuki,R., Tomita,M., Wagneel,L., Washio,T., Sakai,K., Oikio,T., Furuno,H., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boiffin,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Butt,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hochmann,L., Hume,D.A., Kamlay,M., Lee,N.H., Lyons.P., Marchionni,L., Mashima,J., Mazzaelli,J., Momberters,P., Nordone,P., Ring.B., Ringwald,M., Rodriguez,T., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-Oka,K., Wang,K.H., Weitz,C., Whitteker,C., Wilming.L., Wysshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kottski,S. and Hayashizaki,Y.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
JOURNAL	Nature 409 (6821), 685-690 (2001)	
MEDLINE	21085660	
PUBMED	11217851	
REFERENCE	5 (bases 1 to 4001)	
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arikawa,T., Baldarelli,R., Bono,H., Brownstein,M., Butt.C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hangaki,T., Hara,A., Hayatsu,N., Hill.D., Hiramoto,K., Hiroko,T., Hori.F., Hume.D., Imotani,K., Ishii,Y., Itoh.M., Izawa.M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Koya,S., Kutluhan,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa.C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schirml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takahashi,T., Teijima,Y., Toyo,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and	

JOURNAL	TITLE
Hayashizaki,Y. Submitted (10-JUL-2000)	Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp), URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216) Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
	CNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research group in Riken Genomic Sciences Center and Genome Science laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAAGATCCCAAGACCTCTTTTCTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAAGATTCCTCAGTTAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size compressed longer than 7 kb was selected before cloning. Vector: a modified plasmid KSI(+) after bulk excision from lambdaBla FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
	Location/Qualifiers
	1..4001
	/organism="Mus musculus"
	/strain="C57BL/6J"
	/db_xref="FANTOM.DB:4631434016"
	/db_xref="MGD:MGI:1897732"
	/db_xref="taxon:10090"
	/clone="4631434016"
	/tissue_type="skin"
	/clone_lib="riken full-length enriched mouse cDNA library"
	/dev_stage="0 day neonate"
	1..4001
	/gene="Hiplr"
	114..3320
	/note="data source:MGD, source key:MGI:1352504, evidence:ISS huntingtin interacting protein 1 related putative"
	/codon_start=1
	/protein_id="BAB29420.1"
	/db_xref="GI:12852463"
	/db_xref="MGD:MGI:1352504"
	/translation="MNSIKNPVAVLSRTPGHSLAEAREDFDKTOAISKAINSDEAPVEKHARRILILGTHHEKGAFTEWSVAIGLPLSSSTLIISFKCHLVAKYVLDGPANVLHDYORYSNIRSEIDGLMDGHLDYOGHLIVTYTKLLIKIFSHLPQPAFLGTEDVELVLEAGSDVNINIEQLATEMEDVDMDCEKLSESRYRONTAIVASOMSGOCRAPLFLIOCDSSLYHYTYKMLKLSCLPADTLQGRDRPHFOHSIKNFRPARDIMLYFKRLIQIPRDLEGPPNFLRASALAEHTKRPVVIVTEAPEEEPEEPENILEITSPAPGEVYVAADLEDPQFPNGSNKWDNRDLQIEMLKKEVFETLRALKIKMEARYISOLKGVNGILEAELEEFKKOKAKLVNDQELRHLELAQALREARNOGREAEERASXTEARYSKYLKEHSELINHAELLRKFNADYAKOLITOOSOEBAVVKCOLAFOMQARESMKMEKEGDQLEKLRELTAARGELARAQEALESYSESSSLDTLNARKALSGVRRORBAELLAQSIVREKEALSQEORSOKEBELRQALREKSQEBGLQKILLDBDLAVVRSAAAEVALILODAVSKLDDPLHLRCHSSPDVIVSRQAALDSVSGLEGHTVOYLASSADSALVALIAALTRFSHLAADPTVNGAASLAPDPPARLDMFCVGLRECGRALDELVGLODOYVLPAPQSLMKAPLOGIILQGGDLPSLDVROBEICAMVDKEMAATISAILEDVAVRLEDMMSOARHSHSSGVKLFVNERILNSCIDLMKAIRLLWMTSLOKTEIVESRGAAATOOFTFYAKNSRTFEGILSASKAYGMATQVYESADKYVLMKRETELIVCSHETAASTAQVLAESKVRANKNSPMLSRLQECSTRVNEAANAIVASTSGEQIIDRIDTMDSVGLSLAIKIKQMETOVAVRLELEKTLEAEVRRLBELROAHVAVLAGSGMCPSEEPERSPSAPRSGATKKPPAPKQPSIAPRTDNQIDKKDGYYAPQLVNV"
BASE COUNT	925 a 1120 c 1260 g 695 t 1 others
Alignment Scores:	

FEATURES	Class: shotgun.	Location/Qualifiers
SOURCE	1..2191	/organism="Rhodobacter sphaeroides" /strain="2.4.17" /db_xref="taxon:1063" /clone="8758P1D03701098" /clone_1lb="Cosmid library of chromosome II" /lab_host="E. coli S17-1" /note="Vector: pLA2917"
BASE COUNT	341 a 749 c 718 g 382 t	1 others
ORIGIN		
Alignment Scores:		
Pred. No.:	0.284	Length: 2191
Score:	119.00	Matches: 94
Percent Similarity:	37.99%	Conservative: 42
Best Local Similarity:	26.26%	Mismatches: 144
Query Match:	4.93%	Indels: 79
DB:	17	Gaps: 14
US-09-825-414-7 (1-486) x A0012154 (1-2191)		
Qy 143	TripleProaAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla	162
Db 1672	TGCGGGCCGGGGCCCCGCG-AAACCTGCCGACACGCTCTGCATCGCGCTGATCTCTCGCG	161
Qy 163	MetaspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuIstYrLeuSer	182
Db 1613	ATGGAGCAGACCTTGCCCGACACTTTCGCCGGGTTTGGCGGAGGCGTGGCGCGCG	155
Qy 183	AlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArg-----	198
Db 1553	GCACGACGAGATCCGCTTAGTCACACACGCGCCCTCAATCACCGGAGCATGCATCG	149
Qy 199	---HisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTy	217
Db 1493	GCTCATGCTCT-----CGAGTCGAACA--	147
Qy 218	SerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgPro-----	235
Db 1469	-----ACGAGGTGGTGGCTTCAGCTGCGCACCGTCGTCGAGGCGGTGAAC	1422
Qy 236	AlaValGlnGlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAla	255
Db 1424	GCCTCGCTGGCGCCCTCACG-----GGCCTGAGCCGCCGCGT	1386
Qy 256	GlyPheGlyAsnArgLeuSerValGlnSerArgAspHis-----GlnArg	271
Db 1385	GCGGCCACGTCGCGAGATGTCGTAATGCATGGAGAGATTTTCGGCCACGGCGAG	1320
Qy 272	GlyGlyAlaLeuValLeuGly-----LeuLysAspLysGlnPuroLysAlaGln	287
Db 1325	GCCGCTTCGCGGTCACGGGAGTCAGAGCAGGCTCGAGCAGTCGAGCGCGCGCG	1266
Qy 288	LeuSerGlnGluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db 1265	GCCTCGCGAAACCGTGGCGGCGCGCGCCGACGAGATCGCGCGCTCGGAGAAAGTCGT	1206
Qy 308	GlyAlaAlaLeuAsn-AlaGlyLysArgMetAlaGlyLeuProLeuAspMet-----	324
Db 1205	GGAAGGCGCTCCGCTTCATGAGAAAGATCGCGGAACATATTCTGTCATCGAGGAAT	1144
Qy 325	-----AlaThrAspAlaMetGlyAlaValAlaCysLeuValSerAlaSerSerLeuThr	342
Db 1145	CGCGGGGAGACCGATCTTCTC--GCCCGAATGGCGGCGGTGAGAGCCGCCCGCGGG	1088
Qy 342	rgLhAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAl	362
Db 1088	CGACGGGCGCTTCGCTTTCAGTCGCGGCGACGAGTCACGAACATCTCGCGAGGCGAG	1022
Qy 362	aThrLysAsnIleThrAspProAlaThrLysAlaAlaValaSerGlnLeuThrAsnLeuAl	382

Db	1028	C-----CAGGACCGCGCCACAGGA-----GTCCTGACCCCTCTCCACTCGAC	987
Oy	382	aglySerAlaalaValaPhealaGlyTrpThrThraAlaalaLeuthrThraAspProAlaVala 	402
Db	986	GTCGACGACGCGCGCGCGCGGAGATATCATCGGCGCGCTCATCCCGGAT-----	935
Oy	402	LlysAlaAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrl	422
Db	934	-----ATCGAACTCATCTCCACGCTCGCGAC--	908
Oy	422	YTyValAlaAspGlnThrValLysLeuAlaAlaYThrValLysAspMetGlyGlyAla	442
Db	907	-----GACATTCGACATCCGTCGACGCGGAGATCTAGCGCGGAGAT	870
Oy	442	alLeThrHisThrGlyAla-SerLeuArgAsnThrValAsnAsnLeuArgGlnArgPro	462
Db	869	CCGCGCGGTGACCGCTTCGATGCACACGCTCCACGACGCTCAGCAGAGAAATTCCTCCT	810
Oy	462	laArgGlnAlaAspIleGluGlyGlyThrAlaAlaLaserProSerGlu	478
Db	809	ATCGG-----ACGAACCTGCTGGAACGCGCTCTGCCGAGCTCTCAG	772
RESULT 5	5		
LOCUS	BG369810/c	883 bp	linear
DEFINITION	BSVME10026C02f Hordeum vulgare 20 DAP spike EST library HVCNA0010 (20 DAP)	Hordeum vulgare cDNA clone BSVME10026C02f, mRNA sequence.	
ACCESSION	BG369810		
VERSION	BG369810.2	GI:16326017	
KEYWORDS	EST.		
SOURCE	Hordeum vulgare.		
ORGANISM	Hordeum vulgare Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poidea		
REFERENCE	1 (bases 1 to 883) Wing,R., Close,T.J., Kleinbols,A., Wise,R., Begum,D., Frisch,D., Y ,Y., Henry,D., Palmer,M., Rambo,T., Simons,J., Chol,D.W., Fenton ,R.D., Close,S.J., Oates,R. and Main,D. Development of a genetically and physically anchored EST resource for barley genomics: Morex 20 DAP spike cDNA library Unpublished (2001) On Mar 8, 2001 this sequence version replaced gi:13258911.		
TITLE			
JOURNAL			
COMMENT	Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel: 864 656 7288 Fax: 864 656 4293 Email: twing@clemson.edu Total hg bases = 245 Seq primer: AATTACCCCTCAGTAAGG High quality sequence stop: 588.		
FEATURES			
source			
	1..883	/organism="Hordeum vulgare"	
		/cultivar="Morex"	
		/db_xref="taxon:4513"	
		/clone="HVSME10026C02f"	
		/clone_lib="Hordeum vulgare 20 DAP spike EST library	
		HVCNA0010 (20 DAP)"	
		/lissue_type="20 DAP spike"	
		/lab_host="SOLAR"	
		/note="Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, SJ Close, TJ Close). Whole spikes with awns trimmed were collected at 20 DAP (Fenton , Total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give plasmid SK(-) cDNA phagemids in the TJ Close lab at the University of California, Riverside (Choi). Phagemids were plated and picked at the Clemson University genomics institute (CUGI)	


```

Db 1543 TCCATCAGATTGACATGCGATGCAAAATGTTTGTCTCATAGATGATACACGGTGTGGA 1484
QY 121 ATGATAThProPhe-----AlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIle 138
Db 1483 AAGGGGTTCGGCTCGCGCTCAGATGCACTTCAGCTTGAAGCCCGGTTCACGTTGCT 1424
QY 139 AsnLysGlyAspThrLeuProAlaProLeuLysProLeuThrProLeuIle-----Ser 156
Db 1423 TTCACGGGC---TGGCTGTCTCTCTCGCTTCGCCGCTTGTCCGACAGGGTGTCTTACT 1367
QY 157 GAlaAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGly 176
Db 1366 GGAGCGCGCTGTGGCTCG-----GCTCCAGAGC 1340
QY 177 AspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAspSerVal 196
Db 1339 -----GCGGCATCAGTC 1328
QY 197 LysArgHisSerProSerLeuAlaArgGlnValLeuAspThrGlyAlaAlaValGlnThr 216
Db 1328 ----- 1328
QY 217 TyrSerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAla 236
Db 1327 -----GCGGCACACATCTGCTTCGCTTCGCTTCGATCAGCGCGGCA 1283
QY 237 ValGlnGlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGly 256
Db 1282 -----GCATCTGATTTG--ATTTCAGGTGCTGGATTCAGCGGACAGA--GCAGGC 1238
QY 257 PheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuVal 276
Db 1237 TCGGCATCTGATTTGTTTCA-----GCGCTCGAGACAGACA 1199
QY 277 LeuGlyLeuLysAspLysGlnProLysAlaGlnLeuSerGlnLysAsnAspThrLeuGln 296
Db 1198 GCATCAGCAGCAGAT----- 1184
QY 297 AlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArg 316
Db 1183 -----TCAGCACTGTGTTGGTCTCGGCACCTGGACAGCA----- 1148
QY 317 MetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuVal 336
Db 1147 ---GCAGGCTCGCATCTGATTTGGTTTCAGGCGCGGAGACACAGACATCAGACAGA 1091
QY 337 ---SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyIlePheAla----- 353
Db 1090 GATTTCAGCAACGTAGTTTGGTCTCGGGCACTGGAGCAGCAGCAGCTTCACGACCTGAT 1031
QY 354 -----GlyValGlyLysLeuGlnGlnMetAlaThrLysAsnThrLeuThrAspProAla 370
Db 1030 TTGGTGTCTCAGGCGCTGGACAGCAGCAGCTCAGCACTGATTTGGTTCAGGCACTGGA 971
QY 371 ThrLysAlaAlaValSerGlnLeuThrAsnLeu-----AlaGlySerAla 385
Db 970 GCAGCAGCAGCAGCAGGCTCGGCATCTGTTGTTTCAGCGCTGGGACAGACAGCA 911
QY 386 AlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLysAla 405
Db 910 GCAGGCTTGGCATCTGCTTGGTTTCAGCGCTGGGCGGAGACAACACACAGCTTCGGA 851
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrArgGlyTyrValAla 425
Db 850 ACTGATTCGCTCAGCAGCAGCTGGGCGAGACAGCAGCAGCTCGGACAGAGAGGCGCT 791
QY 426 AspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyLysAlaIleThrHis 445
Db 790 GAGTTCTCAGTTGGTTG-----CTTACGTGAGCTGAGCACACAGCT 749
QY 446 ThrGlyAlaSer---LeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGlu 464
Db 748 TCAGGCTCATCTGGCTTGGCATCTTCTTCTCTCC----- 713

```

QY 465 AlaAspIleGlnGlyThrAlaAlaSer 475
Db 712 TCCGATTGCCGCAAGAGAGACTTACACAGCTTCA 680

RESULT 7
BC021514
LOCUS
DEFINITION
Mus musculus, similar to Treacher Collins Franceschetti syndrome 1, homolog, clone IMAGE:5351783, mRNA.
ACCESSION
BC021514
VERSION
BC021514.1 GI:18204412
KEYWORDS
HTC.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK
COMMENT
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-remail.nih.gov

Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@nrl.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stancitrop, S., Thomas, P.D., Touchman, J.W., Tsugeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRK Plate: 55 Row: 1 Column: 23
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6755741
This clone has the following problem: frame shifted.

FEATURES
source
location/Qualifiers

1..4088
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:5351783"
/tissue-type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."
/clone_id="NCI_CGAP_Mam6"
/lad_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 1100 a 1210 c 1210 g 568 t
ORIGIN

Alignment Scores:
Pred. No.: 2.72 Length: 4088
Score: 113.50 Matches: 124
Percent Similarity: 32.40% Conservative: 73
Best Local Similarity: 20.39% Mismatches: 225
Query Match: 4.71% Indels: 186
DB: 11 Gaps: 24

US-09-825-414-7 (1-486) x BC021514 (1-4088)

Qy	11	ProValThrAla----	ThrAspSerPheAlaGlyThrAlaSerAspAlaSerLeuAlaSerSer	29	
Db	1186	CCAGTCACTGTGAATAC	TCTCTCAGCGCAAGCACTTGTGGGAAGCCCTCGGGCCAGAGT	1245	
Qy	30	SerValArgSerValSerSerAsp	IngInArGgUuIleAsnAlaIleAlaAspIleu	49	
Db	1246	ACCTCAGCCCCC	CCAAAGAGTATCCAGAAAGGGGCTCTGCAGTCACCCCCTGGAAAG	1305	
Qy	50	ThrAspAlaValPheAlaAla	HisLysLeuProProAlaAspSerAlaAspIcylInAla	69	
Db	1306	GCAAGGCGCTGTGGAG	CGCCAGCGAGGAACCAAGAACCAAGACAGTGAAG-----	1356	
Qy	70	AlaValAspValHisAsnAla	GlnIleThrAlaLeuIleGluThrArgAlaSerArgLeu	89	
Db	1357	-----	GAQTCAGAGAGTGAAC-----	1371	
Qy	90	HisPheGluGlyGlnThrPro	AlaThrIleAlaAspThrPheAlaLysAlaGluLysLeu	109	
Db	1372	-----	AGTGGGAGACACCAGCTGCT-----	1392	
Qy	110	AspArgLeuAlaThrThrThr	SerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer	129	
Db	1393	-----	-CGCACTGTGACCAAGAGTCCGCAAGAGGTGAACCTTTGGGAAGAGCTCC	1443	
Qy	130	LeuLeuGlnIlyrMetGlnPro	-----	AlaIleAsnLysGlyAspTrpLeuPro	145
Db	1444	CAGGTCAAGACCTGTTC	CCACCCTCACCCGGGCTCATCGGAAAGAGTGCCAACTCTGCC	1503	
Qy	146	AlaProLeuLys-----	-----	ProLeuThrProLeu	154
Db	1504	TGCCCTTGGAGAGGTGGG	GTCAGCACTCTCAGGGGTCMAAATGTAAGAAAGAACAGTGC	1563	
Qy	150	-----	-----	ProLeuThrProLeu	154
Db	1564	TCGGAGAGCAGCAGTGC	AGAGCTGGACAGTGAAGCGGGCTTGGGAGGCCACCAAGCCAAAG	1623	
Qy	155	IleSerGlyAlaLeuSerGly	AlaMetAspGlnValGlyThrLysMet--MetAspArg	173	
Db	1624	GCTCTCCCTTGCCTCT	CTCAGAAAGAGAGCCCTGTGGCCCAAGTCACAGTCAAGACTCAGAG	1683	
Qy	174	AlaThrGlyAspLeuHisThr	LysSerAlaSerProAspArgLeuHisAspAlaMet---	192	
Db	1664	GGCAAGAGCCACTTC	AGGAGCACTGAGAGTCACTGTGACAGCGAAGAGAGGACGACACCA	1743	
Qy	193	AlaAlaSerValLysArg	HisSerProSerLeuAlaArgGlnVal-----	207	
Db	1744	GCACCTCTCTGCTGCT	CAAGGCTTAAGCCAGCTCTGGAAACCAATGAACCTCTCTTGAAG	1803	
Qy	208	-----	-----	LeuAspThrGlyValAlaValGlnThrTySerAlaArgAsnAla	222
Db	1804	AAAGCAGACGCTGCAT	CCGCAACAGAGGAGGAGACACTGCTCCATTGTGAAGCAGAGACG	1863	
Qy	223	Val-----	-----	ArgThrValLeuAlaProAlaLeuAla-----	232
Db	1864	GTGACCTCTTCAAGCA	CCCTGTCACTCCAGACTGTGGCCAAAGGACCCAGAGGTCAAGAT	1923	
Qy	233	-----	-----	SerArgProAlaValGln	238
Db	1924	GTGACACTCTTCCAGT	AGCTGTAGTACAGAGAGCACTGCTCCAGCAACCCCAAGGGGTATAG	1983	
Qy	239	GlyAlaValAspLeuGly	ValSerMetAlaGlyIleuAlaAlaAsnAlaGlyPheGly	258	
Db	1984	GGGAAG-----	-----	TCTGGGGCCAAAGGCTCTCCAAAGGAAAGCTCTTGGGG	2028
Qy	259	AsnArgLeuLeuSerVal	GlnSerArgAspHisArgGlyGlyAlaLeuValLeuGly	278	
Db	2029	CAAGGGGTGGCCCA	AGTCACT-----	CAGAACACAGGGGCTCTGGTCAAACT	2079
Qy	279	LeuLysAspLysGluPro	LysAla-----	GlnLeuSerGluGlnAsnAspTrp	294
Db	2080	ATTGGCTAGGAAGCT	CTCAGAGAGCACTCCAGAGCGAGGAAAGAGATATAG	2139	

QY	295	LeuGluAlaTyr-----	-----LysAlaIleLysSerAla	304
Db	2140	ACCcCAGCAGCAGCCACGCCCTTGGGAGACTTCCTCAGGCCAACCCACCCTCC		2199
QY	305	SerTyrSerGlyAlaIleuAsnAlaGlyLysArgmetAlaGly-----		319
		::: ::::	::: ::::	
Db	2200	ACTAAGACACCTCCAGCGCTTCGATCTGGAAAAAGCTGTGGCTGCTCCAAAGGAAAA		2259
QY	320	LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValAspSerLeuValSerAla---		338
		:	:	
Db	2260	CCACCGTTCGCAACAGCACCGCTCTCGCAAGGGGCCAGCGGTGTGTGCCAGCGCGGGA		2319
QY	339	-----SerSerLeuThrGlnsncLysLeuAlaLeuAlaGlyGlyPheAlaGlyVal		355
		:::	:::	
Db	2320	AAAGCAGGGGCCCCAGCAACCCA-----GCCAGAAAGGGTCCCGTGGCTGGCCA		2370
QY	356	GlyLysLeuGlnGlnMetAlaThrLysAsnIleThrAsp-----ProAla		370
		:	:	
Db	2371	GGGGAGGACTCAGAACGACGACTTAAGAGAGAGTCTGCACAGTGAAGAAAGACGCCACCC		2430
QY	371	ThrLysAlaAlaValAspGlnLeuThrAsnLeuAlaGlySerAlaAlaVal-----		387
		::::	::::	
Db	2431	CAGATAAACCTGTGGGAGACCTCGCAGGTCAGAGCTGCCTCAGCCCTGCCAAGAG		2490
QY	388	-----PheAlaGlyTTPThrAlaAlaLeuThr---		397
Db	2491	TCTCTAAAAAGAGCCCATCCAGAAACCCCGCAGAGCGGATCTTCAGCTTACCAG		2550
QY	398	-----ThrAspPro		400
Db	2551	GCCcCAGCAGGAGAACAGAGACTCAGACAGCACTAGTGAAGACTCGACAGTATACACA		2610
QY	401	AlaValLysLysAlaGlnSerPheIleGlnAspThrValLysSerThrAlaSerSerThr		420
		:::	:::	
Db	2611	GAGATGCCATCAGCCAGAGATGCTATCTCCAGCCCTGCACAGAGCCAAAGCGCAGAGTCCA		2670
QY	421	ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp-----		437
		:	:	
Db	2671	GCGTCC-----CCAGAGAAAGACATAGAAAGGCTCTCAGAGACGACGTAGAGATCTG		2724
QY	438	MetGlyGlnGluAlaIleThrLysThrGlyAlaSerLeuArgAsnThrValAsnAsnLeu		457
		:	:	
Db	2725	CCCTCCGGCAGCAGCATTAATCCCTCCAGTTTCTGTG-----AACCGT		2769
QY	458	ArgGlnArgProAlaArgGluAlaAspIleGlnGlnGlyThrAlaAla-SerProSe		477
Db	2770	AACAGTAGTCCACGCTGTCCACGCTCTCAACCCAGAAAGAGTCCACGCTGTGAACACACA		2829
QY	477	rGluIleProPheArgProMet		484
Db	2830	AAGAAGCCTCAGCACCACTG		2851
RESULT 8				
LOCUS	AY107104	2012 bp	linear	HTC 25-MAY-2002
DEFINITION	Zea mays PCO148976 mRNA sequence.			
ACCESSION	AY107104			
VERSION	AY107104.1	GI:21210182		
KEYWORDS	HTC.			
SOURCE	Zea mays.			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC			
	clade: Panicoideae; Andropogoneae; Zea.			
REFERENCE	1 (bases 1 to 2012)			
AUTHORS	Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,			
	Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.			
TITLE	Maize Mapping Project/DuPont Consensus Sequences for Design of			
	Overgo Probes			
JOURNAL	Unpublished (2002)			
REFERENCE	2 (bases 1 to 2012)			
AUTHORS	Coe,E.C.			
TITLE	Direct Submission			

JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
 FEATURES Location/Qualifiers
 source 1. 2012
 /organism="Zea mays"
 /db_xref="MaizeDB:638791"
 /db_xref="taxon:4577"
 /clone="PC0148976"
 /library="Maize Mapping Project/Dupont Consensus
 Library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed Dupont contigs; this resource was
 assembled by Dupont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

BASE COUNT 463 a 621 c 597 g 331 t
 ORIGIN

Alignment Scores:

Pred. No.: 1.18 Length: 2012
 Score: 112.50 Matches: 82
 Percent Similarity: 33.33% Conservative: 51
 Best Local Similarity: 20.55% Mismatches: 174
 Query Match: 4.66% Indels: 92
 DB: 11 Gaps: 9

US-09-825-414-7 (1-486) x AY107104 (1-2012)

QY 8 GlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAla 27
 Db 648 CAGAAAGCAGCCGCTGTCGCGGCTACGTCGCGCAGCAGCAGATGATAGGCTCGG 707
 QY 28 SerSerSerValArgSerValSer-----SerAspGlnArgGluLeuAla 44
 Db 708 CTCGCCAGCGCGCGCCGAGATGATGCTTCGACATCAAGGCCAGTCTTTCACGA 767
 QY 45 IleAlaAspTryLeuThrAspHisValPheAlaAlaHis-LysLeuProProAlaAspSe 64
 Db 768 GCAGAACCTTCACATTTGACATGATGCATGACATGCGGCTCGCATG----- 822
 QY 64 PAlaAspGlyGlnAlaAlaValaAspValHisAsnAlaGlnIleThrAlaLeuIleGluTh 84
 Db 823 -----TCCGTTCTCAGCGCTCAGCTTCGTTCTCTTC 860
 QY 84 rArgAlaSerArgLeuHisPheGluGluThr---ProAlaThrIleAlaAspThrPh 103
 Db 861 CTCGGCGGGGCACTCTCTATGACAAACACCTGTCGCAACGGGTGATCTCTCC 920
 QY 103 eAlaAlaGluGlyLeuAspArgLeuAlaThrThrThrSerGlyAla-LeuAlaGAlaT 123
 Db 921 GAAGGCGAGCGGCGCTCTGCTCCATGCGCAAGGCGCCCTTCAMTGGCGTATGACGAT 980
 QY 123 hrProPheAlaMetAlaSerLeuGlnTryMetGlnProAlaIleAsnLysGlyAspT 143
 Db 981 CTCGCCAGGCGAGCTGATCAGCAACAGCGGCAACGCAAGCTTACCTCTGACATT 1040
 QY 143 rPLeuProAlaProLeuLysProLeuThrProLeuIle-SerGlyAlaLeuSerGlyAla 162
 Db 1041 GG-----GTTCTTCTGCGCAACAGTGGTGGCAGC 1070
 QY 163 MetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTryLeuSer 182
 Db 1071 TCTGCGAGCCAGGCGCTCTGCTCCAGCAAGAGCCCTTCAMTGGCGTATGACG 1130
 QY 183 AlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSer 202
 Db 1131 CAACCTCCCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1190
 QY 203 LeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTrySerAlaArgAsnAla 222
 Db 1191 CTTGGG-----GTTCTTCTGCGCAACAGCGGTGGATGGCCCGGAGCAGCAGCATCAGGAC 1244

QY 223 ValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValaAsp 242
 Db 1245 GCGGCGCTGTGATCCACAGACCACTC----- 1271
 QY 243 LeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeu 262
 Db 1272 -----AACGGCGGCGAGCAGCAGCAACCGCAGCTGATTAACGGGATGGCG 1316
 QY 263 SerValGlnSerArgAspHisGlnArgGlyGlyValaLeuValaLeuGlyLysAspLys 282
 Db 1317 TCCATCGGAGC-----CACTTGGCGGAGCGCTTCGTCCTTCCTGACACACTCA 1367
 QY 283 GluProLysAlaGlnLeuSerGlnGluAsnAspPrpLeuGluAlaTyrLysAlaIleLys 302
 Db 1368 TCTCCGTCGCGCGGCGCTCA----- 1388
 QY 303 SerAlaSerTrySerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu 322
 Db 1389 -----GGCAGCGCGCTC 1400
 QY 323 AsPheAlaThrAspAlaMetGlyAlaValaArgSerLeuValSerAlaSerSerLeuThr 342
 Db 1401 CTGATGAAGCGGCTCAGATGGCTTCGACGACGTCGACACACCAACCCACAGCGC 1460
 QY 343 GlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValaGlyLysLeuGlnGluMetAla 362
 Db 1461 CCCAGCGCGTGTCTCCGCGCAGCGCGCTTCGACGACACTGCGCAGCGGAGACTAGT 1520
 QY 363 ThrLysAsnIleThrAsp----- 368
 Db 1521 AGCAGGCGACGGAGAGAGAACTCTACAGGCAATATTCATGATTCATCATGAAAC 1580
 QY 369 -----ProAlaThrLysAlaAlaValaSerGlnLeuThrAsnLeuAlaGly 383
 Db 1581 TCACTGCCATTAAAGGAGGAGGACCGCGGCTCTCAGGCTTGGGGCG 1631
 RESULT 9
 AL545051 975 bp mRNA linear EST 16-FEB-2001
 LOCUS AL545051 L7L.NFL006.PL2 Homo sapiens cDNA clone CS0D1013YF04 5
 DEFINITION
 AL545051
 AL545051.1 GI:12877532
 VERSION
 KEYWORDS
 EST.
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 975)
 Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 JOURNAL
 Unpublished (2001)
 COMMENT
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 Location/Qualifiers
 source 1. 975
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="CS0D1013YF04"
 /clone_lib="L7L.NFL006.PL2"
 /tissue_type="placenta"
 /note="Vector: pCMVSPORT 6; Site: 1; NotI: 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-stranded cDNA was digested with Not I and
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6
 vector. Library was normalized. Library was constructed by
 Life Technologies. Contact : Feng Liang Life Technologies,
 a division of Invitrogen 9800 Medical Center Drive
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
 Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com"


```

Db      930  TCCGGACATCTGTCAGGTTGGAACTTACGACCTGCTGCTCCCTAGATCCCA 871
Oy      233  -----SerArgProAlaValGlnGlyAlaValAspLeuGlyAlaSerMet 247
Db      870  AGGCTCTCTGGACGACAGCGGCTGTGTATTTGGGCGACCTAACACTGGGAATCAACC 811
Oy      248  AlaGlyGlyLeuAlaAlaAsnAlaGlyPhe-----GlyAsnArgLeuLeu 262
Db      810  CCCACTGGATGGACAGATCTCTTGGGGTCTGTGTCACATTGCTGCTTCCCCCATGGCG 751
Oy      263  SerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeuGly----- 278
Db      750  GAAGCCCGGAAGCGATCTCATCGATGAGAGTCCGCCATTGATGTCAGAACTTTGAAGTAG 691
Oy      279  LeuLysAspLysGluProLysAlaGln-LeuSerGluGlnAsnAsp----- 293
Db      650  GTGGAGATGAAAGAAAGACAGAACGACGACAAAGTGAAGTGAAGACGTAGGACAC 631
Oy      294  -----TrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAl 309
Db      630  ACAGTTGCTCCACATACGTAAGACACAT----- 603
Oy      309  AlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAla 329
Db      602  -GCCCAATGGAATTTGAGGTCC-----AAGTTGGAAGAGCCCTGCACACGCAAT 553
Oy      329  tGlyAlaValArgSerLeuValSerAlaSerSerLeuProGlnAsn---GlyLeuAla 348
Db      552  GGCACCTGGACCTGGACCCCTGCTGACAGCTTCCACGACGATGAACCATGGGATG- 498
Oy      348  uAlaGlyGlyPheAlaGlyAlaGlyLysLeuGlnGlnMetAlaThrLysAsnIleThrAs 368
Db      497  -----GGGCCAGACCCACTTCAAGAAAGGCCACAAAG----- 465
Oy      368  ProAlaThrLysAlaAlaVal-SerGlnLeuThrAsnLeuAlaGlySerAlaAlaVal 388
Db      464  ---CCAAAGATGGCCAGACAGTCTCAGATAGCAGATCCAGGGTAGCTGCTCCAGCAGTGC 409
Oy      388  heAlaGlyTrpThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerP 408
Db      408  TAGCCGATGTCATGATGATGCTA-----CAACCCGCCATGCCAGCAGAGCCT---- 359
Oy      408  heIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGln 428
Db      358  --ATGAGGTGACGAGGTCCGCGCTGCTCCATCCACCAAC-----ACGACAGCA 307
Oy      428  hrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHisThrGly- 447
Db      306  CAGTGAAGCGCCGTGTGACGATACCGAG-----CCAATGGTGGCATACACAGGCT 256
Oy      448  -----Ala-SerLeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArg 463
Db      255  GCTGCACCCCGCCTTCTCGAAGATGCTGCTGGAGTAATAGACAGCGTGTGATGCCAG 196
Oy      464  GluAlaAspIleGluGlnGlyTyrThrAlaAlaSerPro 476
Db      195  ACAGCT-----GCTGGACAGCTGCACGACCA 169

```

```

RESULT 11
LOCUS   BM588321 711 bp mRNA linear EST 25-FEB-2002
DEFINITION 1700687322518 A.Gam.ad.cdna blood1 Anopheles gambiae cdna clone
19600449697628 5', mRNA sequence.
ACCESSION BM588321.1 GI:18884182
VERSION   BM588321.1
KEYWORDS  EST.
SOURCE    African malaria mosquito.
ORGANISM  Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 711)

```

```

AUTHORS Holt,R.A., Lin,J.-J., Murphy,S.D., Evans,C.A., Kraft,C.L., Charlab
TITLE    R., Collins,F.H., Venter,J.C. and Hoffman,S.L.
JOURNAL  Celera Anopheles gambiae EST project
COMMENT  Unpublished (2002)
Contact: Holt R.A.
Celera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: Holt@celera.com
Plate: N001004AAR row: F column: 14
Seq primer: M13 Reverse.
FEATURES
source
1..711
/oranism="Anopheles gambiae"
/strain="RSP-ST (Reduced susc. to Permethrin - std.
chromosome)"
/db_xref="taxon:7165"
/clone="19600449697628"
/clone_1lb="A.Gam.ad.cdna.blood1"
/dev_stage="Adult"
/lab_host="MDH10b"
/note="Vector: pSport1; Site_1: SalI; Site_2: NotI; whole
adult mosquitoes (mixed sex) frozen on liquid nitrogen 24
hours after human blood feeding. cdna inserts >500 bp
cloned directionally into pSport 1. Not 1 site is 3'.
Clones available through the Malaria Research and
Reference Reagent Resource Center (www.malaria.mr4.org)"
BASE COUNT 305 a 234 c 135 g 37 t
ORIGIN
Alignment Scores:
Pred. No.: 0.396 Length: 711
Score: 110.50 Matches: 56
Percent Similarity: 33.19% Conservative: 23
Best Local Similarity: 23.53% Mismatches: 144
Query Match: 4.58% Indels: 15
Gaps: 5
US-09-825-414-7 (1-486) x BM588321 (1-711)
Oy      218  SerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaVal 237
Db      37  ACAGCAGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 96
Oy      238  GlnGlyAlaValAspLeuGlnValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPhe 257
Db      97  GCACACGACACGACACCAACAATTCGACGACGACATCAACGACACCAATATGCACAGAT 156
Oy      238  GlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeu 277
Db      157  GCATCAAGAAATGACATCTGCTCAACAAAACGACGACATCGTGCACATCAACAA 216
Oy      278  -----GlyLeuLysAspLysGluProLysAlaGlnLeuSerGluGlnAsnAspThrLeu 295
Db      217  TGTCCAGGA--AAATCAACAGCATCTGCAGCATGCAACGATCAATCAATCAATTT 273
Oy      296  GluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLys 315
Db      274  ACAACATGACGACGACATCAAGCAACAAATATCGACACACGACAAACACCAACACGCA 333
Oy      316  ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeu 335
Db      334  GCCACAGCGCC-----GCCACAGCCAGAACCTCAGCTGCGACGACACGCA 378
Oy      336  ValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyPheAlaGlyAl 355
Db      379  ACAACAGCAGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 438
Oy      356  GlyLysLeuGlnGlnMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaVal 375
Db      439  GCAGCAACAGATCAACAGCAACGACGACGACGACGACGACGACGACGACGACGACGAC 498

```

OY	376	sergineLeuthrasleuAlagIySerseValalaValPhealagIyTPrThlValala	395
Db	499	GCAGCACACGACAACAGCAGCAACGACACACCGACACACACAGCAGCA---ACACACACACCA	555
OY	396	LeuthrThrAspProAlaValLysLysAlaGluSerPheilegInAspThrValLysSer	415
Db	556	ACAAACAGCAGCATGCACAAACAGCAGATGCA-----ACAGAACACACGA	597
OY	416	ThrAlaSerSerThrThrglyTYrValAlaLaspglnThrValLysLysAlaThrVal	435
Db	598	GCAGCCGGCAACAAACAGCACTGCACAGCACACAGCCGCCGACGACGACGCCGCCCA	657
OY	436	LysAspmetGlyGlyGUAlaIleThrHisThrGlyAlaSerLeuArgAsnThr	453
Db	658	GCAACAAACAACAACATCTGCACAGCAACAGCAGCAGCAGCAGCAGCAACAACAACA	711
RESULT 12			
LOCUS	CNS076CP/c	888 bp	DNA linear GSS 07-JUL-2001
DEFINITION	T7 end of clone BB00A002G01 of library BB00A from strain CBS 4732		
ACCESSION	Pichia angusta, genomic survey sequence.		
VERSION	AL431231		
KEYWORDS	AL431231.1 GI:12214643		
SOURCE	GSS.		
ORGANISM	Pichia angusta.		
REFERENCE	Eukaryote; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Pichia.		
	1 (bases 1 to 888)		
	Souciat,J., Aigle,M., Artiguenave,F., Blandin,G.,		
	Bolotin-Fukuhara,M., Bon,E., Brotier,P., Casaregola,S.,		
	de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,		
	Malpertuy,A., Neuvéglise,C., Ozler-Kalogopoulos,O., Potier,S.,		
	Saurin,W., Tekata,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,		
	Wincker,P. and Weissenbach,J.		
	Genomic exploration of the hemiascomycetous yeasts : 1. A set of		
	yeast species for molecular evolution studies		
	FEBS Lett. 487 (1), 3-12 (2000)		
TITLE	2 (bases 1 to 888)		
JOURNAL	Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.		
MEDLINE	and Dujon,B.		
PUBMED	Genomic exploration of the hemiascomycetous yeasts : 13. Pichia		
REFERENCE	angusta		
AUTHORS	FEBS Lett. 487 (1), 76-81 (2000)		
	20584723		
	1152888		
	3 (bases 1 to 888)		
TITLE	Genoscope.		
JOURNAL	Direct Submission		
MEDLINE	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,		
PUBMED	2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :		
REFERENCE	seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
AUTHORS	This GSS is part of a random genomic sequencing program of thirteen		
TITLE	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces		
JOURNAL	exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,		
	Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces		
	lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia		
	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,		
	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to		
	5 kb were prepared and both extremities were sequenced. See		
	keywords for description of this sequence and for the sequence of		
	the other extremity of this insert.		
COMMENT	Location/Qualifiers		
	1..888		
FEATURES	/organism="Pichia angusta"		
SOURCE	/strain="CBS 4732"		
	/db_xref="taxon:4905"		
	/clone="BB00A002G01"		
	/clone_1lb="BB00A"		
	/note="end : T7"		
BASE COUNT	225 a 190 c 273 g 194 t 6 others		

[illegible]

Qy 6 ArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSer 25
Db 74 CAATTGAAGCAAGACCAGAGTAGATGCGTCTAATCCGATGCTCCAGC 133
Qy 26 LeuAlaSerSerValArg-----SerValSerSerSp 37
Db 134 CTGACATGGGATTCCTCCAGACGACAGATCTGCAGATGACACCGAAGGAGCTAT 193
Qy 38 GlnGlnArgGluLeuAlaAlaAlaAspTyr-----LeuThrAspHisValPhe 54
Db 194 CCAGGACGAGCAGAGTGTCTCCCTACAGACATAGAAATGGAGGCAACAGCCAT---TTT 250
Qy 55 AlaAlaHisLysLeuProProAlaAspSerAlaAspGlnAlaAlaValAlaValHis 74
Db 251 AACTTACTGGCCTTCCTCCGACACGACGCTCCCTCAGGACGAAACCC----- 298
Qy 75 AsnAlaGlnIleThrAlaLeuIleGluThrArgLaserArgLeuHisPheGluGly 94
Db 299 TCTCCTTCCTGGGGGAGGCGCTACACTAACGGGTCTCCCATGAACTCCCGCAGCA 358
Qy 95 ThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAla 113
Db 359 GGGAAAGTTTGATGGGAT-----GTGAATGTTAATGGCTTATCTACTGATCTAC 412
Qy 114 ThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer----- 129
Db 413 ACTACTACTCAGGAGATTTGAACTCTCCGACCTCTAGACCTCAGACCTCCAT 472
Qy 130 -----LeuLeuGlnIlyMetGln---ProAlaIleHisLys 140
Db 473 CACCTTAACGTGGCCCTACGACTGTCTTGGAACACTACAGTACCCATGCCATCT 532
Qy 141 GlyAspThrLeu-----ProAlaProLeuLysProLeu 151
Db 533 GGCACAAACCTCCAGGAGCCACCCCTTCTTCTCAGTCCCTGGGGGCAATACCCGCTC 592
Qy 152 ThrProLeuIleSerGlyAla----- 158
Db 593 AACGGTACCTTGGGGGCAACCCGACACCTTCATCCCACTGACACACATACTTCTGA 652
Qy 159 -----LeuSerGlyAlaMetAspGlnValGluThrLysMetMet 171
Db 653 GCTGGAGCCCAAGAGTGTGGGCCAATGTATGCCAGAGTCCCAATGGGCTTAATCTGAT 712
Qy 172 AspArgAlaThrGlyAspLeuHisTyrIleuSerAlaSerProAspArgLeuHisAspAla 191
Db 713 TCACAGGAACGTATGAT-----TCTTTCCGATCAACAATTTTGAGAGTG 757
Qy 192 MetAla-----AlaSerValLysArgHisSerProSerLeuAlaArg 205
Db 758 ATGGCCCAATGGACCCCAAGTTTTCACCTCCCTCCAGCTCTCCCAATGTG----- 811
Qy 206 GlnValLeuAspThrGlnValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThr 225
Db 812 -----GGGTCTAGTATCCAGACCTTTGCACTTCCCGAGATGATGACAGT 856
Qy 226 ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal---AspLeuGly 244
Db 857 GACATCCATCCTGATGAGAGCAGCAAAAGAGAGTACTGATGTTGTCGCAAAATAGGC 916
Qy 245 ValSerMetAlaGlyGlyLeuAlaAlaAsn-----AlaGly 256
Db 917 ACTGGCTTGTGAGCAGCTGAGCTGAGAGAGAGCAGACAGAACTAAAGATGTGTGC 976
Qy 257 PheGlnAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuVal 276
Db 977 TACATGCTGTCTCTCTCTGTGTGAGTCTTTA---CACCAAGAGTCTCCGCTCTGCTC 1033
Qy 277 -----LeuGlyLeuLysAspLys---GluPro 284
Db 1034 CCTGATCCACAGTAGAGCTGTCTAGATGATCCTTCACATCTTCCGATCAACAGTGAAGAC 1093
Qy 285 LysAlaGlnLeuSerGluGlnAsnAspThrLeuGlnAlaIleThrLysAlaIleLysSerAla 304

Db 1094 ACTCCAAATTCACATGAA-----GACTCCCTGGAGCCCTTTGACTCTCTGGCAGCAG 1147
Qy 305 SerTyrSerGlyAla-----AlaLeuAsnAlaGlyLysArgMetAlaGly----- 319
Db 1148 CCACTGAGTGGAGCTTTTATGTATGATGATGCGGAGCTGATGGGTGCAGAAAGACAG 1207
Qy 320 LeuProLeuAsp----- 323
Db 1208 TTGCTCTGGAGGCAACCCGTGATCTCTGCCCTCATTCCTGCTCCGATATGCT 1267
Qy 324 -----MetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla 338
Db 1268 AATGCTTCACTCTCTGGCAGACAGACAGACATCAGCCTCCATCTTGTGAGCCCT 1327
Qy 339 SerSer-----LeuThrGlnAsnGlyLeuAlaLeuAlaGly 350
Db 1328 ACCTCCCCACCTGTCTAGGGAGCTGTCTTGCAGATACAGTACACTTTGGACTG----- 1381
Qy 351 GlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsn----- 365
Db 1382 AACAGTTCAGTCACTCTGAACAGAGAAATAGAGACCCAGTCTTCAAACTTCCAACT 1441
Qy 366 ---IleThrAsp-----ProAlaThrLys 372
Db 1442 CCCCCTGACTGAGCAGAGCTCTGACAGCCAGCCTAGTACTCACTACATCAGACGATTTCA 1501
Qy 373 AlaAlaValSerGlnLeuThrAspLeuAlaGlySerAlaAlaValPheAlaGlyTrpThr 392
Db 1502 CCACAGCCTCCCGACAGAGCCTCTCTGACAGACATCTGCAGAAATCTCTCCAGCTGTCT 1561
Qy 393 ThrAlaAlaLeuThrThrAspPro-----AlaValLysLysAlaGluSer 407
Db 1562 CCAGTAGATCTCGCGCTGCTCCGTAAGTCTTTGAGCAGCTCTCCAGCTTCTCA 1621
Qy 408 PheIleGlnAspThrValLysSerThrAlaSerSerThrGlyLys----- 423
Db 1622 CCTGCTCTGCCAGCATCTCTTTGGAAGCCTGTGACACACTCCAGTAACCTTCTCTCA 1681
Qy 424 -----ValAlaAspGlnThrValLysLeuAlaLysThrVal 435
Db 1682 GGTTCCTCCAGACTCTTCCAGCAGCTGTCTCCAGACTGTCTCCAGCAGAG----- 1735
Qy 436 LysAspMetGlyGlyAlaAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsn 455
Db 1736 -----AAA 1738
Qy 456 AsnLeuArgGlnArgProAlaArgGluAlaAspIleGlu-----GlyGlyThrAla 473
Db 1739 AATGTCAGAGCTGCTCTAAAGCGCGTGTGATGAGAGAGAGAGAGCTGGAGGAGAGTT 1798
Qy 474 Ala 474
Db 1799 GCA 1801

Search completed: January 31, 2003, 07:16:15
Job time : 1595.38 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:12:39 ; Search time 54.444 Seconds
(without alignments) 2737.582 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412

Sequence: 1 MHNRVQGPVTATDSFRT.....IEEGTAAPEIFRRMRS 486

Scoring table:

BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 883724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_epool/US09825414/runat_28012003_161742_261144/app_query.fasta.1.1294
-DB=Issued_Patents_NA -QEMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414 -ECGN_1_1_40 -crunat_28012003_161742_26114 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAR -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV=TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: Issued Patents NA:*
- 2: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCRTUS.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	7.0	4403765	4	US-09-103-840A-2
2	164	6.8	4411529	4	US-09-103-840A-1
3	147.5	6.1	6744	1	US-08-119-125A-2
4	143.5	5.9	45819	4	US-09-453-702B-72
5	138.5	5.7	4411529	4	US-09-103-840A-1
6	137.5	5.7	32768	4	US-08-961-527-71
7	135.5	5.6	2550	4	US-08-961-527-364
8	133.5	5.5	5036	4	US-09-177-349-2
9	132.5	5.5	4403765	4	US-09-103-840A-2
10	130.5	5.4	44377	2	US-08-804-227C-7
11	130.5	5.4	44377	2	US-08-804-198-1
12	129	5.3	4483	4	US-08-961-527-363

C	13	128.5	5.3	38506	3	US-09-320-878-19	Sequence 19, Appl
	14	128	5.3	16836	4	US-09-147-236-1	Sequence 1, Appl
	15	128	5.3	16836	4	US-09-147-236-10	Sequence 10, Appl
	16	127.5	5.3	3106	2	US-08-591-079-5	Sequence 5, Appl
C	17	127.5	5.3	11220	4	US-09-105-537-32	Sequence 32, Appl
	18	127.5	5.3	15872	4	US-09-105-537-1	Sequence 1, Appl
	19	127.5	5.3	36778	4	US-09-105-537-5	Sequence 5, Appl
C	20	127.5	5.3	50341	1	US-08-247-901C-1	Sequence 1, Appl
	21	127.5	5.3	50341	2	US-09-075-904-1	Sequence 1, Appl
	22	127.5	5.3	52297	4	US-09-426-436-1	Sequence 1, Appl
	23	127.5	5.3	52297	4	US-08-705-557-1	Sequence 1, Appl
	24	127	5.3	3251	4	US-09-085-199B-6	Sequence 6, Appl
	25	127	5.3	3300	1	US-08-194-290-6	Sequence 6, Appl
	26	127	5.3	3300	2	US-08-614-377A-6	Sequence 6, Appl
	27	127	5.3	3300	4	US-09-142-648B-6	Sequence 6, Appl
	28	126.5	5.2	9542	4	US-08-968-685A-9	Sequence 9, Appl
	29	126.5	5.2	2406	1	US-07-689-008-4	Sequence 4, Appl
	30	126.5	5.2	4977	4	US-09-071-035-257	Sequence 257, App
	31	126.5	5.2	4977	4	US-09-071-035-261	Sequence 261, App
	32	126.5	5.2	4977	4	US-09-071-035-265	Sequence 265, App
	33	126.5	5.2	9540	1	US-07-689-008-1	Sequence 1, Appl
	34	126	5.2	1702	4	US-08-961-527-367	Sequence 367, App
	35	125.5	5.2	3979	4	US-09-085-199B-10	Sequence 10, Appl
C	36	123.5	5.1	8078	4	US-09-702-251-3	Sequence 3, Appl
	37	123	5.1	6530	2	US-08-146-930-1	Sequence 1, Appl
C	38	123	5.1	6530	5	PCT-US93-03993-1	Sequence 1, Appl
C	39	123	5.1	28958	1	US-08-258-261B-6	Sequence 1, Appl
	40	122.5	5.1	28958	1	US-08-456-837-6	Sequence 6, Appl
	41	122.5	5.1	28958	1	US-08-457-342-6	Sequence 6, Appl
	42	122.5	5.1	28958	1	US-08-457-646A-6	Sequence 6, Appl
	43	122.5	5.1	28958	1	US-08-458-076A-6	Sequence 6, Appl
	44	122.5	5.1	28958	1	US-08-764-233A-4	Sequence 4, Appl
	45	122.5	5.1	28958	1		

ALIGNMENTS

RESULT 1

US-09-103-840A-2/C
Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Alignment Scores:

Pred. No.: 0.0596

Score: 168.00

Percent Similarity: 37.67%

Best Local Similarity: 26.02%

Query Match: 6.97%

DB: 4

US-09-825-414-7 (1-486) x US-09-103-840A-2 (1-4403765)

Oy 145 ProAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMet 164
 Db 3936672 CCGGCGCG-----CGGCTGCCCGCTGGTACTAGAGAGCGTGGCGGCGCTCCG 3936619
 Oy 164 spGlnValGlyThrLysMetLeuAspArgAlaThrGlyAspLeuHisIstLysSerAlas 184
 Db 3936618 GTGGCGCGGACACCGGCTTGGCCCGCTGCCACCGT----- 3936581
 Oy 184 erProAspArgLeuHisAspAlaMetAlaValAsaSerValLysArgHisSerProSerLeuA 204
 Db 3936580 -----GCCGTATGGCGCGGACCGCCGCGCGGCGGACCGCGCG----- 3936545
 Oy 204 IaArgGlnValLeuAspThrGlyValAlaValAlaGlnThrTyrSerAlaArgAsnAlaValA 224
 Db 3936544 -----CGCCCGACCGGACACCC----- 3936527
 Oy 224 rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValaLeu 244
 Db 3936526 -----GCGCGCGCTTGACCTCTTGGCCGCGGTGGCGCGGACGTGGTGG 3936478
 Oy 244 lYValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg----- 260
 Db 3936477 GCGCCGCT---GCGCGCGCGCGCGCGCGGACCGCGGACCTTGGCGGTCCGCGCGTGGCG 3936421
 Oy 261 --LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaValaLeuValLeuGlyLeuL 280
 Db 3936420 CCTTTGCCCGCGCGGACCGACCCGATACCGCGCGCGCGGCGGCGCACACT----- 3936371
 Oy 280 yAspLysGlnProLysAlaGlnLeuSerGlnLysAsnAspThrLeuGlnAlaTyrLysA 300
 Db 3936371 ----- 3936371
 Oy 300 IaIleYSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL 320
 Db 3936370 -----GCCGCTCTGGCGCGCGGTGCCCGCTGGCGCGCGGACCGCGGTGCCGCG 3936319
 Oy 320 euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAlas 339
 Db 3936318 CCACCGTTGCTGGACCGCGGCGGCGCGCGGACCGCGCGGTGGCGCGGTGCCGCGCG 3936259
 Oy 339 erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly---GlyPheAlaGlyValGlyLysL 358
 Db 3936258 GTACCGTTGGCGCGCGGACGATACCGCGCTGGTGGCGCGGACACCGGACCGCGCGCTT 3936199
 Oy 358 euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL 378
 Db 3936198 CCGGCCCTTCCGCGACT-----CCACCGCTTCCGCGCGCTATTTCGCGCTA 3936154
 Oy 378 euThrAsnLeuAlaGlySerAlaAlaValaPheAlaGlyTyrThrThrAlaAlaLeuThrT 398
 Db 3936153 CCACCGGCGCTTCGCGGACGCGCGCCGATGGCGCGCTTGGCGCGGTGGCGCGGTACT 3936094
 Oy 398 hrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr----- 416
 Db 3936093 CCGGTCCGCGCGCGGACCGCGGTGCATCCGCTACCGGACGACCGCGGTTTCCACCGTTG 3936034
 Oy 417 -----AlaSerSerThrThrGlyTyrValA 425
 Db 3936033 CCGCGCGGCGCGCGGCTTCCGCCATGGCCCGCGCGCGGACGCGCGGTTGGCGCGGAGATG 3935974
 Oy 425 IaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle----- 443
 Db 3935973 CCGACGCGACACGCGCGGCATCTCCGCGTGGCGCGGCAATTCGACAGGATACGTTGGCG 3935914
 Oy 444 -----ThrHisThrLysAlaSerLeuAlaGlnThrValaLysAsnLeuArgGlnArgProA 462
 Db 3935913 CCGGAGATGGACACGCGGCGCGC-----GCGCGCGCGCTCTCCGCTTCCGCGCGCTGGCG 3935857
 Oy 462 IaArgGlnAlaAspLleGlnGly-----GlyThrAlaAlaSerP 476
 Db 3935856 CCGCGCGCGCGCGCTTGGCGGCGGTCTGTGTATACCGCTTGTCCACCGTGGCGCGCT 3935797
 Oy 476 roSerGlnIleProPheArgPro 483

DB	3935796	TTGGCCGCGATTGCGCGCCGCC	3935774
DB	3935796	TTGGCCGCGATTGCGCGCCGCC	3935774
RESULT 2	US-09-103-840A-1/c		
	Sequene 1, Application US/09103840A		
	Patent No. 6294328		
	GENERAL INFORMATION:		
	APPLICANT: FLEISCHMAN, Robert D.		
	APPLICANT: WHITE, Owen R.		
	APPLICANT: FRASER, Claire M.		
	APPLICANT: VENTER, John C.		
	TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM		
	TITLE OF INVENTION: TUBERCULOSIS		
	FILE REFERENCE: 24366-20007.00		
	CURRENT APPLICATION NUMBER: US/09/103,840A		
	CURRENT FILING DATE: 1998-06-24		
	NUMBER OF SEQ ID NOS: 2		
	SOFTWARE: PatentIn Ver. 2.1		
	SEQ ID NO 1		
	LENGTH: 4411529		
	TYPE: DNA		
	ORGANISM: Mycobacterium tuberculosis		
	OTHER INFORMATION: H37Kv		
	US-09-103-840A-1		
Alignment Scores:			
Pred. No:	0.153	Length:	4411529
Score:	164.00	Matches:	95
Percent Similarity:	37.40%	Conservative:	43
Best Local Similarity:	25.75%	Mismatches:	133
Query Match:	6.80%	Indels:	98
DB:	4	Gaps:	14
US-09-825-414-7 (1-486) x US-09-103-840A-1 (1-4411529)			
OY	145	ProAlaProLeuLysProLeuThrProLeuLysSer--GlyAlaLeuSerGlyAlaMetA	164
DB	3943125	CCGGCGCGCG-----CCGCTCCGCGCGCTGGTACTGAGGAGCGCTGGCGCGCGCGCG	3943072
OY	164	SPGlnValGlyThrIysMetMetAspArgAlaThrGlyAlaSerLysIleuSerAlaAs	184
		:::	
DB	3943071	GTGGCGCGGACCGCGCTGGCCCGCGTGGCCACCGT-----	3943034
OY	184	erProAspArgIeuHisAspAlaMetAlaAlaSerValIysArgHisSerProSerLeuA	204
DB	3943033	-----GCCGTGACCGCGCCACCGCGCGGCCGCCACGCCCGC-----	3942998
OY	204	IaArgGlnValIeuAspThrGlyValAlaValGlnThrIysSerAlaArgAsnAlaValA	224
DB	3942997	-----CGCCCGACCGCGCACAC-----	3942980
OY	224	rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuG	244
DB	3942979	-----GCGCGCGCGCTTGACCTTCCTGGCCCGCGGTGGCCCGCGGACTGGTGTG	3942931
OY	244	lyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg-----	260
DB	3942930	GGCCCGCT---GCCCGCGCGCGCGCGCGCGCACCGCGCGCACTTCGCCGCGCGCTTCCG	3942874
OY	261	--LeuIeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuL	280
DB	3942873	CCCTTGGCGCGCGCGGACCGACCGATACCGTCGCGCGCGCGGCCACACACT-----	3942824
OY	280	ysAspLysGluProLysAlaGlnIeuSerGluGlnAsnAspTriLeuGlnAlaTrpLysA	300
DB	3942824	-----3942824	
OY	300	IalleLysSerAlaSerIyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL	320
DB	3942823	-----GCCGCTTCGCGCGCGCTCCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCG	3942772
OY	320	eupProLeuAspMetAlaThrAspAlaMetGly---AlaValArgSerLeuValSerAlaAs	339

Db 3942271 CCACGGTTCGTCACCGCGTGGCGGCCCGCGAGCGCCCGCTTCGCGCGCGCCG 3942712
QY 339 erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly--glyPheAlaGlyValGlyLysL 338
Db 3942271 GTCACGGTTCGCGCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3942656
QY 358 euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL 378
Db 3942655 -----GTTCCCGCCCTTCCCGCCCAATCCACCGCTTCCCGCGCTATTCGCGCTA 3942607
QY 378 euThrAsnLeuAlaGlySerAlaAlaValAlaAlaGlyTrpThrAlaAlaLeuThrT 398
Db 3942606 CCACCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3942547
QY 398 hrAspProAlaValLysLysAlaGlySerPheIleGlnAspThrValLysSerThr--- 416
Db 3942546 CCGGTCGCTG 3942487
QY 417 -----AlaSerSerThrThrGlyTyrValA 425
Db 3942486 CCGAGATG 3942427
QY 425 laAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIle--- 443
Db 3942426 CCGAGCGGACACGCT 3942367
QY 444 -----ThrHsrThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA 462
Db 3942366 CCGGAGGTGGCACCG 3942310
QY 462 laArgGluAlaAspIleGluGluGly-----GlyThrAlaAlaSerP 476
Db 3942309 CCGCT 3942250
QY 476 roSerGluLeuPropheArgPro 483
Db 3942249 TTGCGCGCATTCGCT 3942227

RESULT 3
US-08-119-125A-2
; Sequence 2, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, poly
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the diagn
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Central Diergeneeskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Leystad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v 6.0
; SOFTWARE: Wordperfect v. 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,125A
; FILING DATE: 20-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00054
; FILING DATE: 19-MAR-1992
; APPLICATION NUMBER: NL 9100510
; FILING DATE: 21-MAR-1991
; ATTORNEY/AGENT INFORMATION:

NAME: Handal, Anthony H.
REGISTRATION NUMBER: 26275
REFERENCE/DOCKET NUMBER: SMITHHE119125
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 838-8589
TELEFAX: (203) 838-8794
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 6744 base pairs
TYPE: Nucleic acid with corresponding amino acids
STRANDEDNESS: single stranded
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptococcus suis type II (pathogenic)
FEATURE:
OTHER INFORMATION: Extracellular factor related protein (EF*) gene
FEATURE:
NAME/KEY: promoter -35 region
LOCATION: bp 66 to 71
FEATURE:
NAME/KEY: promoter -10 region
LOCATION: bp 89 to 94
FEATURE:
NAME/KEY: promoter -35 region
LOCATION: bp 153 to 158
FEATURE:
NAME/KEY: promoter -10 region
LOCATION: bp 176 to 181
FEATURE:
NAME/KEY: ribosome binding site
LOCATION: bp 350 to 356
FEATURE:
NAME/KEY: signal peptide
LOCATION: bp 361 to 498
FEATURE:
NAME/KEY: start of repetitive units R1-R11
LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
5065, 5293, 5521:
FEATURE:
NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences
LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
5128, 5356, 5584:
FEATURE:
NAME/KEY: dyad symmetry regions
LOCATION: from bp 6534 to 6566 and from bp 6571 to 6583
FEATURE:
NAME/KEY: dyad symmetry regions
LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644
US-08-119-125A-2

Alignment Scores:
Pred. No.: 0.000199 Length: 6744
Score: 147.50 Matches: 105
Percent Similarity: 36.76% Conservative: 81
Best Local Similarity: 20.75% Mismatches: 219
Query Match: 6.12% Indels: 101
DB: 1 Gaps: 19

US-09-825-414-7 (1-486) x US-08-119-125A-2 (1-6744)

QY 14 AlAtThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSer 33
Db 4024 GCTAGAGTCCAGTTCGATTGCTAGGATTAAGAAATTAAGGAAATCCGAAACA 4083
QY 34 ValSerSerAspGlnArgGluIleAsnAlaIleAlaAspPyrLeuThrAspHisVal 53
Db 4084 GAAGAAGAGAGACCTACTAAATAGTAGAGAAATCTGCAGAA-----GATACGCCG 4134
QY 54 PheAlaAlaHisLysLeuProAlaAspSerAlaAspGlyGlnAlaAlaVal 73
Db 4135 AAGGCTATGAGAGACAAATCCAACTTGTCCAGTAGAAGATTAAGCAGCGAAATTAAG 4194

```

OY 74 HisAsn-----AlaGlnIleThrAlaLeuIleGluThrArgAlaSerArgLeuHis 90
Db 4195 CTAACAGAGCTGTGGCAAACTTGAACACCATCTGACAACTGACAACTGACAGTAC 4254
OY 91 PheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAsp 110
Db 4255 CAAAGAAAGCAAAAGAACTCAAGCCCTAAGCAGAT--CTTGAAAGAAAGTAAAGCAAG 4311
OY 111 ArgLeuAlaThrThrThrSer-----GlyAlaLeu 120
Db 4312 AAAATTGCAGATAAAGCTCGATGTAGTGTACTACTTGTGAAGATGTGAGCTT 4371
OY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 4372 GAAGCTACT-----AAACAGATGCTAAGACAAAG 4401
OY 141 -----GlyAspTyrLeuPro 145
Db 4402 ATTGCTAAAGATGCAAGCCGCTGCTAAAGACAAATTGCAAGCAATTCGAACAGAC 4461
OY 146 AlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db 4462 GCAGAGAAAGAAACCTTCACCGATGGCGTAGATGCAGAGTAGTCCAAAGCTAAGCAGCA 4521
OY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
Db 4522 ATTTCAGCT-----GCACCAGC-----CCAGCAGATGTT 4551
OY 186 AspArgLeuHisAspAlaMetAlaIleAlaSerValLysArgHisSerProSerLeuAlaArg 205
Db 4552 CAAAGAAAGAGAGATGTCAGCTGTGACGCAATTCGCAAGAAAGTGTTCGACGAGCTAAG 4611
OY 206 GlnValLeuAspThrGlyAlaValAlaGlnThrTyrSerAlaArgAsnAlaValArgThr 225
Db 4612 CAAGATGCTAAGAAATGATGCTTAAAGATGCAGCCGCTGCTAAGACGAATGCTCTCC 4671
OY 226 ValLeuAlaProAlaLeuAlaSerArgProAlaValAlaGlnGlyAlaValAspLeuGlyAl 245
Db 4672 AATCCAAACTTGAACAGCCGCAAGAAACCTTCACCGATGGCTAAGTGAAGCA 4731
OY 246 SerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGln 265
Db 4732 GCCAAAGCTAAGACGCAATTCAGCTGCA-----ACCAGCCACAGATGTTCAA 4782
OY 266 SerArgAspHisGlnArgGlyGlyAlaLeu-----ValLeuGlyLeuLysAspLys 282
Db 4783 AAAGAAGAGAGATGCAAGCTGTGACGCAATTCGCAAGATGCTTTCGACGAGCTAAGCA 4842
OY 283 GluProLysAlaGlnLeuSerGluLysAsnAspTyrPheGluAlaTyrLysAlaIleLys 302
Db 4843 GATGCTAAGAAATGATGCTTAAAGAAATCCGAC-----GCTGCTAAG 4884
OY 303 SerAla-----SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMet 317
Db 4885 TCAGGCATTGAGCGCAATTCGAACCTTGAAGATGCAGAGAAAGAAATCAGCTAAGAAAGCA 4944
OY 318 AlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSer 337
Db 4945 GTTGATGCTGATGCTAAAGCTGCGACAGATGCAATGATGCTTCAACAGTCAGTCGAA 5004
OY 338 AlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyAlaGlyLys 357
Db 5005 GCGCAATCGCGCAAGACAAA-----GCGCTAGGTTCA 5037
OY 358 LeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGln 377
Db 5038 ATC-----GCCAAGATGTTCTTGACGCGCAAGCAAGATGCTAAGAAACAAAG 5085
OY 378 LeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrAlaAlaLeuThr 397
Db 5086 ATTGCCAAAGAAAGTTCGCGCAGCT-----AAAGACGAATGTAT 5124
OY 398 ThrAspProAlaValLysLysAlaGlu-----SerPheIleGlnAspThrValLysSerThr 416

```

```

Db 5125 GCCAATCCCAACTTTCAGATGCAGAGGAAGCTTCTAAGAAAGCGGTATGATCAGAT 5184
OY 417 AlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu---AlaLysThrVal 435
Db 5185 GCTAAAGCTTACACAGATGCAATGATGCTTCAACAAAGTCAGTGAACGCCAATCGGCA 5244
OY 436 LysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsn 455
Db 5245 GAGGACAAAGCC-----GTAGGTTCAATGCGCCAAAGATGTTCTTGAC 5286
OY 455 AsnLeuArgGln-----ArgProAlaArgGluAlaAspIleGluGluGly 471
Db 5287 GCAGCGAAACAAGATGTAAGATTAAGATGCTAAGAAATCGCAGCTGCTAAGTCAGCC 5346
OY 472 ThrAlaAlaSerProSer 477
Db 5347 ATTGACGCAATCCAAAC 5364

RESULT 4
US-09-458-702B-72
; Sequence 72, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatner, Frederick R.
;            Perna, Nicole T.
;            Plunkett, Guy
;            Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plunkney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-9166
; TELEFAX: (608) 251-5000
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46819
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-453-702B-72

Alignment Scores:
Pred. No.: 0.0118 Length: 46819
Score: 143.50 Matches: 107
Percent Similarity: 33.40% Conservative: 52
Best Local Similarity: 22.48% Mismatches: 232
Query Match: 5.95% Indels: 85

```

DB: 4 Gaps: 15
US-09-825-414-7 (1-486) x US-09-453-702B-72 (1-46819)
QY 19 ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln 38
DB 32015 CGTACGCGGCTCGCGGTGGACAGAACGCGCGGCAAGACGATGACGATGCC 32074
QY 39 GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaHisLys 58
DB 32075 AGCAGATACGCCCGTAGCGGCGGACACCAATGCGACTGAT-----GCTCCGAGCTCA 32125
QY 59 LeuProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle 78
DB 32126 GCAGCGGAGCAGCAGCAGCTCAGCCGCGAGCGCGGTGCTGCGCTCAGCAGCGCTTCC 32185
QY 79 ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGlyGluThrProAlaThr 98
DB 32186 AGCGCA-----GGACGCGCATCAACAAAGGCTACT 32215
QY 99 IleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGly 118
DB 32216 GAGGATCAAAAAGTGTGCGCTCAGAGTCTCAAAAAGCGCGGCTACAGTGC 32275
QY 119 AlaLeuArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnProAlaIle 138
DB 32276 GGTGGCGGAAACGCTCAGAAACGAAATGCGGAGTGTACAA-----CAATCAGCGCGC 32329
QY 139 AsnLysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAla 158
DB 32330 ACTTGTGCA-----TCCACCGCGGACCAAGCAAGCTCAGAGCT 32368
QY 159 LeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrLysPheLeu 178
DB 32369 GCCTCTCAGCGGAGGATGCGTGCCTCAAAAAGCGCGCAAAATCATCA--GAAACG 32425
QY 179 HisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAla-----SerVal 196
DB 32426 AGCGAGCGCTCGAGCGCCAGTGTGACAGCTCTCGGACAGCGCGGCAATTCGCG 32485
QY 197 Lys-----ArgHisSerProSerLeuAlaArgGlnVal 207
DB 32486 AAGCGCGCAAAACGCTGAGACAAACGCTAAGCTCTGAAACGCGCAGACAGAGCT 32545
QY 208 LeuAspThrGlyValAlaValGlnThrTyrSerAla-----ArgAsnAlaValArgThr 225
DB 32546 GCCTCCGACGACGAGCGCTCAAAAACGCGCTGCATTATCTGCCAGTCCCGCTCAACA 32605
QY 226 ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyVal 245
DB 32606 AGTGGCGGCGAGCGCTCAGCCAGCTGCCACCGCGCGGCAAAATCGGACAGAAAGTCCGCA 32665
QY 246 SerMetAlaGlyGlyLeuAlaAlaHisAlaGlyPheGlyAsnArgLeuLeuSerValGln 265
DB 32666 TCGTCTGCTTCAACAGCCACAACGAAAGCTGGCGAAGCCACTGACAGAGCGCGCAGACA 32725
QY 266 SerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
DB 32726 CGGAGTCTGCTCCGACGCGAAGACATCCGAACGAAACGCGGATGGAAACCAAC 32785
QY 286 AlaGlnLeuSerGluGlnAsnAspTyrLeuGluAlaTyrLysAlaIleLysSerAlaSer 305
DB 32786 GCAGAAATCTCAAAAACGCGTCCGATCTCAGCCAGTTCGCGCGGCGTCAATCGGATCA 32845
QY 306 TyrSerGlyAlaAla-----LeuAsnAlaGlyLysArgMetAla 318
DB 32846 TCTGCGTCTGCTCAAAAGATGACGACACAAAGCTGACGACGAAAGACGCGCC 32905
QY 319 GlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla 338
DB 32906 ACGAGCGCATCCAGACGAGCGCAGAGCGAGCTGTAGTGGCAGCGACGACCTCAGAGC 32965
QY 339 SerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeu 358

DB 32966 AAAAGTACGCGCGAATCTGCACGACGCGCGCT----- 32998
QY 359 GlnGluMetAlaThrLysAsnIleThrAspProAla----- 370
DB 32999 ---GAGACGCGCAAAAGGCGGACAGAGATATATGATTCGCGCGCTTGAGGATGCG 33055
QY 371 -----ThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPhe 388
DB 33056 AACACGACGCAAAAGGGATATACAGCTCAGCAGTGGCACTAACAC----- 33103
QY 389 AlaGlyTyrThrThrAlaAlaLeuThrThrAspPro---AlaValLysLysAlaGlnSer 407
DB 33104 -----ACTCCGAGTCACTGGCGGCGGACGCGCAAAAGCGGTAAAGCGCGATAGAG 33154
QY 408 Phe-----IleGlnAspThrValLys----- 414
DB 33155 CTGGCTAACGGGAATATACCCGACAGATGCGACAGACAGAAAGGATAGTTTCAG 33214
QY 415 ---SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLys 433
DB 33215 CTTAGCAACGCGCAACAGCAGATCTGAATATGCTGGCGGCAACGCGCAAAAGCGGTAAAG 33274
QY 434 ThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSer 449
DB 33275 GCAGCGTATGACCTTGCTTACGCGGAATATATACCTGCTCAGAGAGCTTACG 33322
RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 60.7 Length: 4411529
Score: 138.50 Matches: 122
Percent Similarity: 33.91% Conservative: 53
Best Local Similarity: 23.64% Mismatches: 193
Query Match: 5.74% Indels: 148
DB: 4 Gaps: 21
US-09-825-414-7 (1-486) x US-09-103-840A-1 (1-4411529)
QY 20 ThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSer----- 35
DB 1632232 ACCGCGCTGCGCGCGCGCGCGGCTTACCGCTTACCGCGCTGACCGCGCTTACCACTC 1632291
QY 36 SerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAla 55
DB 1632292 GCCGAAAGCGAAATGCGCGTGGCGCGCTGCGCGCGCTCAGCAGCGGAGCCCGCGCGCC 1632351
QY 56 AlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsn 75
DB 1632352 CCTTGGCGCGCGGACCGCGCGACACCTGTATTCGTTTGGCCAAAGAGTTCCCGCGC 1632411
QY 76 -----AlaGlnIleThrAlaLeuIleGluThrArgAlaSerArg 88

Db 1632412 CAACCGCCGGGCGCCGCTTGGCCGCCTTAAGCCTTG-----CGCGCC----- 1632456
OY LeuHisPhegluglyGlnThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLys 108
Db 1632457 -----GGGCCCGCCGTGGCCGCGCTGCCTTCCTTTGGCGCCGG----- 1632495
OY IeuAsnArgLeuAlaThrThrThrseryglyAlaIeuArgAlaThrProPheAlaMetAla 128
Db 1632496 -----CGTGTGGCGCTTTGGCGCCGCTTGGCGCGCTTGCGCGCGCCACCAGCGTCCGCC 1632549
OY SerIeuLeuGlnTyrrMetGlnProAlaIleasnLysGlyAspTrpLeuProAlaProleu 148
Db 1632550 GTCCGCCCGCAAGATCTCCGGCCCGCGCC-----GGCGGCCCGCCG 1632588
OY LysProleuthrProleuIleSerGlyAlaIeuSerglyAlaMetaspGlnValGlyThr 168
Db 1632589 CGCCCCCGCACCCCGCGGTTCGGCCGCTCGTCGCGCATTCGCCCGCGGTGGGG-- 1632645
OY LysMetMetAspArgAlaThrGlyAspIeuHisThrIleuSerAlaSerProAspArgLeu 188
Db 1632646 -----CGAGGAACCGGC----- 1632657
OY HisAspAlaMetAlaAlaSerValLysArghisSerProserIeuAlaArgGlnValIeu 208
Db 1632657 ----- 1632657
OY AspThrGlyValAlaValGlnThrTyrrSerAlaArgAsnAlaValArgThrValIeuAla 228
Db 1632658 GACACCGCGCCATGCGTCCCGCTTGGCCGC-----GGC 1632693
OY ProAlaIeuAlaSerArghProAlaValGlnGly-----Ala 240
Db 1632694 CCGCGCATTAACAACAAGCCCGCGGTGGCGCCCTTGCGCGCGCCCGCCCGCGCCCGCC 1632753
OY ValAspLeuGlyValSer-----MetaGlyGlyIleuAlaAlaAsnAlaGlyPheGly 258
Db 1632754 GGCGACGGTGGCGTTCGGCGCGCGCTGGCGCGCGGTGCCCGCGTGGCGCGCGTGGGGGT 1632813
OY AsnArgrLeuLeuSerValGlnSerArghAspHisGlnArghGlyGlyAlaIeuValIeuGly 278
Db 1632814 GCGCGCGGGGACACGCTGTGCACCCCGCGGTGGATTC---GGCGCGCGCATCCACACAGC 1632870
OY IeuLysAspLysGlnProLysAlaGlnIeuSerGlnGlnIuAsnAspTrpLeuGlnIuLar 298
Db 1632871 ACCACCGATGCCGCGCGCTACCGCGCTTGGCGCGCTTGCCACACATCCATCCGCGCGTT 1632930
OY ATCGACACTGCCCTTGGCACCGCTTGGCGCGCATACCGCCCATTCGCCGCGCGCGCGCTT 1632990
OY AsnAlaGly-----LysAlaIleLysSerAlaSerTyrrSerglyAlaIalaIeu 311
Db 1632991 TCCCGCGGCGCCCGCGGACCCATGTCGCGCTCCGTGGGTGGGTGCGAAGCGCCTTACC 1633050
OY GlyLeuProIeuAspMetAlaThrThrspalameGtGlyAlaValArgSerIeuValSerAla 338
Db 1633051 GCCCTTGGCCACCGGCTCCACCGCCACCGCGCGGTCCACCGTTCGCGCGCTTGGCGCGCGTC 1633110
OY SerSerIeuThrGlnAsnGlyIeuAlaIeuAlaGlyGlyPheAlaGlyValGlyLysIeu 358
Db 1633111 GGTCCCATTCGCGCTGCCCCACGACCGCGTAAAGCCGCGTGGCGCGGT----- 1633158
OY GlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnIeu 378
Db 1633159 -----GGCGCGCGTTCGCCCGCGCTTGGCGCCCTT 1633185
OY ThrAsnLeuAlaGlySerAlaAlaValAlaPheAla---GlyTyrrThrThrAlaIaleuThr 397
Db 1633186 ACCCGCGGCGCGCGCAGACGCGCGTGGCGTGGTGGCGTCCGCGCGCGCGCTTACC 1633245
OY ThrAspProAlaValLysLysAlaGluSerPhe-----IleGlnAspThr 412

[illegible]

```

QY      21  AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnGlnArg 40
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4541 TCAGCCTCAGCATCAACAGAGTGGCTTCAGCAAGACTAGTGCATCAGCTTCAACA 4600
QY      41  GlnIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaIleHisLysLeuPro 60
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4601 AGTACTAGAGCGCTCAGCCTCAGCGTGCAGCAAGCGCTCAGCTTCAACA----- 4648
QY      61  ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4649 ---AGTACCAGTGGCTCAGCCTCAGCGTGCAGCAAGTGGCTGGCTTCAACAAGTACCTCA 4705
QY      81  LeuIleGlnThrArgAlaSerArgLeuHisPheGlnGlyGlnThrProAlaThrIle---- 99
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4706 GCGTCTGAATCAGCATCAACAGAGTGGCTTCAGCATCAACGAGTGCATCAGCTTCA 4765
QY      100 AlaAspThrPheAlaLysAlaGlnLysLeuAspArgLeuAlaThrThrThrSerGlyAla 119
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4766 GCATCAACAAGTGGCTTCAAGTCA-----GCAAGTACCAAGTGGCTGGCT 4810
QY      120 LeuAlaGlnThrProPheAlaMetAlaSerLeuLeuGlnTyrTyrGlnProAlaIleAsn 139
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4811 TCAGCATCAACAGTGGCTTCAAGTGCAGTCAACCAAGTGGCTTGAATCCGATCAACA 4870
QY      140 LysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuLysSerGlyAlaLeu 159
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4871 AGT-----GCTGGCTTCAGCAAGCAGCAGTGGCTTGGCTGGCTGCTCA 4915
QY      160 SerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHis 179
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4916 AGGAGTGGCTGCATCAGCATCAACAGAGTGGCTCAGCAGTCAAGC----- 4963
QY      180 TyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHis 199
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      4964 ---ACATCAGCTTCTGAATCTGCATCAACAGTGGCTCAGCTTCCCATCAACAACAGCC 5020
QY      200 SerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaGlnThrThrSerAla 219
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5021 TGGGCTCAGCAAGT-----ACAAAGTCTTCAGCCTCAGCATCAACAGCAGT 5065
QY      220 ArgAsnAlaValaArgThrValLeuAlaProAlaLeuAlaSerArgProAla---ValGln 238
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5066 GCATCAGCTTCAAGCATCAACAGAGTGGCTTCAAGTGCAGTCAACCAAGTGGCTGGCTTCA 5125
QY      239 GlyAlaValaAspLeuGlyValaSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGly 258
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5126 GCAAGTACCAGTGGCTCAGCTTCAAGCAACAGAGTGGCTCAGCTTCAAGCATCAACAGCT 5185
QY      259 Asn-ArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGln 278
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5186 GCTTGGCTTGGCATCAACAAGTGGCTCAGCATCAGCT----- 5225
QY      278 TyrLeuLysAspLysGlnProLysAlaGlnLeuSerGlnGlnAsnAspTyrLeuGlnAlaTyr 298
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5226 -CAACGAGTGGCTCAGCCTCAGCAAGTACTAGTGCATCAGATCAG-----CA 5272
QY      298 rIysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla 318
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5273 TCAAACCAAGTGA-TCAGCTCAGCAAGTATCTCAGCGCTGAATCGGCATCAACGAGTGC 5331
QY      318 acLysLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValaLargSerLeuValSerAla 338
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5332 ATCAGCATCAGCATCAACAGAGTGCATCGCTTCAAGCTCAACCAAGTGCATCAGCTTCAAGC 5391
QY      338 aser---SerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValaGlyLys 357
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5392 AAGCAACCAAGTGGCTGGCTCAGCATCAACGAGTGGCTCAGCCTCAGCAAGTACTAGTGCAGC 5451
QY      357 sIeuGlnGlnMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValaSerGln 377
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5452 GTCCTGAATCGCATCAACAAGTGGCTCAGCCTCAGCAAGTACTAGTGCATCGGCTTCAAGC 5511

```

```

QY      377 nLeuThrAsnLeuAlaGlySerAlaAlaValaPheAlaGlyTyrThrThrAlaIleLeuThr 397
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5512 AACACCAAGTGGCTGGCTTCAAGCATCAACAGTGGCTCA---GCCTCAGCAAGTATCTC 5568
QY      397 rThrAspProAlaValaLysAlaGlnSerPheIleGlnAspThrValLysSerThrAla 417
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5569 AGCGTGTGAATCGCATCAACAGAGTGCATCAGCCTCAGCAAGTACTAGTGCATCAGCCTC 5628
QY      417 aserSerThrThrGlyTyrValaAlaAspGlnThrValLysLeuAlaLysThrValLysAsn 437
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5629 AGCATCAACGAGTGCATCGGCTTCAAGCAAGTACCAAGTGGCTTCAAGCAAGCAGCAG 5688
QY      437 pMetGlyGlyGlnAlaIleThrHisThrGlnAlaSerLeuArgAsnThr-ValaAsnIle 457
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5689 TGGCTACGCTCAGCAAGTACCAAGTGGCTTCAAGTGGCTTCAAGCAAGTGGCTTCAAGC 5745
QY      457 euArgGlnArgProAlaArgGlnAlaAspIleGlnGln 469
      : : : : : | | | | | : : : : : | | | | | : : : : :
Db      5746 TTCAGCAAGTACCAAGTGGCTCAGCTCAGCATCAACAAG 5783

```

RESULT 7

US-08-961-527-364

Sequence 364, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READEABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

CLASSIFICATION: 424

FILING DATE:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 364:

SEQUENCE CHARACTERISTICS:

LENGTH: 2550 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-364

Alignment Scores:

Pred. No.: 0.000686 Length: 2550

Percent: 135.50 Matches: 99

Score Similarity: 37.02% Conservative: 85

Best Local Similarity: 19.92% Mismatches: 259

Query Match: 5.62% Indels: 55

DB: 4 Gaps: 11

US-09-825-414-7 (1-486) x US-08-961-527-364 (1-2550)

QY 2 HisIleAsnArgArgValaGlnGlnProValThrAlaThrAspSerPheArgThrAla 21

```
Db 815 CATCGAATCAGCGT-----CAACAGAT-GCTCGGCTTCAGCAAGTACAGTCCCTCA 867
Oy 22 SerAlaSerLeuAlaSerSerValArgSerValSerSerAlaValInArgL 41
Db 868 GCTTACAGATCAACCAAGCGCTCGCTCAGCAAGACCTCTGAAATCGGCGCTCA 927
Oy 42 IleAsnAlaIleAlaAspTyrIleThrAspHisValPheAlaAlaHisLysLeuPro 61
Db 928 ACCAGCGCTCGCTCAGCAAGCAAGCTTCGATCGGCTCAACCAAGCGCTCA 987
Oy 62 AlaAspSerAlaAspGlyAlaAlaValAspValHisAsnIleGlnIleThrAlaLeu 81
Db 988 GCTCAGATCAACGAGTGGCTTCGCTTCAGCAAGCAAGCGCTCGGCTCAGCATCA 1047
Oy 82 IleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAsp 101
Db 1048 ACGAATAGCTCAGCTTCAGCGCTCAACCAAGTCTTCAGCTCAGCATCAACAGTGGCTCA 1107
Oy 102 ThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeuArg 121
Db 1108 GCCTCAGCAAGATCTCAGCGCTCGCTGAAATCGGATCAAGAGTGGCTGAGTCAAGATCA 1167
Oy 122 AlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGly 141
Db 1168 ACGAATAGCTCAGCTTCAGCAAGCAAGTGGCTTCAGCTCAGCAAGTCTTCAGCATCT 1227
Oy 142 AspThrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGly 161
Db 1228 GAA-----TCGGCATCAACGAGTGGCTCGCTTCAGCAAGTACTAGC 1269
Oy 162 AlaMetAspGlnValGlyThrLysMetLeuAspArgAlaThrGlyAspLeuHisTyrIleu 181
Db 1270 GCCTCAGCATCAGCTCAACAGTCTCGGCTTCAGCGCTCA-----ACG 1314
Oy 182 SerAlaSerProAspArgLeuHisAspAlaMetAlaIleSerValLysArgHisSerPro 201
Db 1315 AGTGGCTCT-----GAGTCAGCATCAACGAGTCAAGTCAAGCTCAGCTCAGCAAGC 1359
Oy 202 SerLeuAlaArgGlnValLeuAspThrGlyValAlaValAlaGlnThrTyrSerAlaArgAsn 221
Db 1360 ACATCAGCTTCGAAATCTGCATCAACCAAGTGGCTCAGCTCAGCATCAACAGCGCTCA 1419
Oy 222 AlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAla---ValGlnAla 240
Db 1420 GCTTCAGCAAGTACAGCGCTCAGCTCAGCAAGTACAGTCAAGCTTCAGCGCTCAGCGTGC 1479
Oy 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaIleAsnAlaGlyPheGlyAsnArg 260
Db 1480 ACAAGTGGCTCGCTCAACCAAGTGCATCTGAATCGGATCAACCAAGTGGCTCAGCTCAG 1539
Oy 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 1540 CAAGTACTAGTCACTCAGCTCAGCATCAGCAAGTGCATCGGCT----- 1584
Oy 281 AspLysGluProLysAlaGln-LeuSerGluGlnLysAsnAspTyrLeuGluAlaTyrLysAl 300
Db 1585 CGGCTCAACCGATCAGTACAGTACAGCAAG-----TACCAGTG 1623
Oy 300 AlLeuLysSerAlaSerTyrSerGly-----AlaAlaLeuAsnAlaGln 314
Db 1624 CGTCACTTCGCGATCAACAGTGGCTCGGCTTCAGCAAGCAACATCAGCATCTGAATCAAC 1683
Oy 314 LysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSer 334
Db 1684 GTCAACCAAGTGGCTTCAGCAAGTACAGTGGCTTCAGCTTCAGCATCAACCAAGCGC 1743
Oy 334 IleuValSerAlaSer---SerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyIlePheAl 353
Db 1744 CTCGCGCTCAGCAAGCACTCAGTCTGAAATCGGCTCAACCAAGCGCTCGGCTCAGC 1803
Oy 353 AglyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAl 373
```

```
Db 1804 AAGCACTCAGCTTCGATGCGGCTCAACCAAGCGCTCAGCTCAGCATCAACAGTGC 1863
Oy 373 alaValSerGluLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrTh 393
Db 1864 TTGGGCTTCAGCAACCAAGCGCTCGGCTTCAGCATCAACAGAT-----ACGTC 1914
Oy 393 talAlaLeuThrTh-----AspProAl 401
Db 1915 AGCTTCAGCGCTCAACCAAGTGGCTTCAGCTCAGCATCAACAGTGGCTCAGCAAG 1974
Oy 401 aValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerThrTh 421
Db 1975 TATTCAGCGCTTCGATGCGGCTCAACCAAGTGGCTTCAGCATCAACCAAGTGGCTCAG 2034
Oy 421 rGlyThrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 441
Db 2035 AGCTTCAGCAACCAAGCTCAGCTTCGATGCGGCTTCAGCTCAGCATCAACAGTGGCTCAGCATC 2094
Oy 441 valIleThrHisThrGlyAlaSerLeuArgAsnThrValAlaAsnLeuAlaArgGlnArgPr 461
Db 2095 GACAAGCGCTCAGCTTCAGCAAGTACCAAGTGGCTTCAGCTCAGCGCTCAGCAAGTGGCTC 2154
Oy 461 oAlaArgGlnAlaAspIleGluGlyGlyThrAlaAlaSerProSer 477
Db 2155 GCGCTCAACCAAGTGCATCTGAATCGGCTCAACCAAGTGGCTCAGCTCA 2203

RESULT 8
US-09-177-349-2
; Sequence 2, Application US/09177349
; Patent No. 6268201
; GENERAL INFORMATION:
; APPLICANT: Alland, David
; APPLICANT: Bloom, Barry R.
; APPLICANT: Jacobs Jr., William R.
; TITLE OF INVENTION: Inb, Inla AND Inlc GENES OF MYCOBACTERIA AND METHODS
; FILE REFERENCE: 96700/491
; CURRENT APPLICATION NUMBER: US/09/177,349
; CURRENT FILING DATE: 1998-10-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; TYPE: DNA
; LENGTH: 5036
; ORGANISM: Mycobacterium tuberculosis
US-09-177-349-2

Alignment Scores:
Pred. No.: 0.00331 Length: 5036
Score: 133.50 Matches: 108
Percent Similarity: 33.41% Conserved: 43
Best Local Similarity: 23.89% Mismatches: 182
Query Match: 5.53% Indels: 120
Gaps: 16

US-09-825-414-7 (1-486) x US-09-177-349-2 (1-5036)
Oy 64 SerAlaAspGlyGlnAlaValAspValHisAsnAla----- 76
Db 97 ACCGTGATGGCAGCGCGGCTCAGCGCGGCAACAGCTTTCGATGGAAGGAATG 156
Oy 77 GlnIleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrPro 96
Db 157 AAGATGACCTCGCTTATGATATACATCTGAGACCTTCGCGGAGGAGAGACCGCCGG 216
Oy 97 AlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThr 116
Db 217 TCGTTCGTTGCC-----GCTCCGGGAGCGGCGCATGTGACGT 252
Oy 117 SerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnPro 136
Db 253 GCGGCGGTGATGATATGCGCGCGCACCAATCTCATGGTGGCGGCAATGTGCTGCCG 312
```

QY	137	AlaLeuSlnuSgLIysAsprIrrPleuProAlaProLeuLysProLeuThProLeuIIeSer	156
Db	313	GGTGTGATGGGTGGCCGGCAGCC	339
QY	157	GlyAlaLeuSerLIyAlaMetAspInValIGlyThrLys-MetMetAspAlaThrIgl	176
Db	340	-----ATGAGCGGATGGCGGAGCGCGCGCGCGCGTTCGGCATGGCTTTGGCCAGAGCTC	393
QY	176	YAspLeuHISLysrLeuSerAlaSerProAspArgLeuHISAspAlaMetAlaAspVa	196
Db	394	GSCA-----ATGCGCGCTTCGGCGGTGACGCGGGCGGGCGGGGTGGCAAGCGTCATC	444
QY	196	LYsArgHISLeuSerProSer-----LeuAlaArgInValLeuAspThrIglValAlaVa	214
Db	445	ACGACCGATGTCCGTGGCGCGCGCTGTACCGGACGTGGGTCTGGGTCTGGGTCAAGGT	504
QY	214	IGlnThrTySerAlaArgAspAlaValAlaArgHr-----	225
Db	505	GGCGTGGCTGTCCCGCGCGCTCAGCGGTGTTCGGCGCGTCAAGTGGCTTGGTGGCCAG	564
QY	226	-----ValLeuAlaProAlaLeuAlaLeuAspArgProAlaValIGlnIlyAlaAla	242
Db	565	GTCGCTGTGGGTTTAACTGGCGGTGATG-----AGCGCGAG-GTCCGCGGTCAAGTTGG	617
QY	242	pleuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLe	262
Db	618	TGCTGGGTTGGATATGGGCGACGCTGGGTGTCTCAGCGCGGTATGAGGC-----	666
QY	262	uSerValIGlnSerArgAspHisGlnArgIglGlyAlaLeuValLeuGlyLeuLysAspLy	282
Db	667	-----TTTGGCGCGGGGTGGCTTGGCTTGGGTGGGTGT--	699
QY	282	SgLIProLysAlaGlnLeuSerGlnLysAsnAspThrPleuGlnAlaTyLysAlaIleLy	302
Db	700	-----CAGCGCGCGGGTGGATCGG-----	719
QY	302	sSerAlaSerTySerGlyValAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLe	322
Db	720	TGGAGCGCGCGCGGGGTATCGCTGCCGCGCGGTGGCTCCCTTACGAGCGGCAATGGC--	777
QY	322	uAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuVal-----	336
Db	778	-CAGATCGAGTTGCCGGCCAGGCGTCCGTGGCGGTGGTGTGGCGCTGGCTGTGCGCG	836
QY	337	-SerAlaSerSerLeuThrGlnAsnGlnLysLeuAlaLeuAlaGlyIglPheAlaGlyValG	356
Db	837	CCAGCGCGGCGATGGCATGCCAGATCGGTGTCTCAGCCGGTGGTGGCGCGCGCGCTGG	896
QY	356	YLYs-----	357
Db	897	CAATGTACGCGCGCTGACGCGGCTCAGCAGAACCGAGTGTGGCTTCAACGCAAGCGG	956
QY	358	-----LeuGlnGlnMetAlaThrLysAsnII	366
Db	957	CCAGCGCGGGTTATGCGCCAGTGAAGCGCTCCCTTGAACGGCGGTGTATATGCTCATCT	1016
QY	366	ethAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAl	386
Db	1017	GTCGGCGCGCTTATGCGCGGTGTGGGTGGTGTGACGCGCGCGCGCTGGCGCGCGCGG	1076
QY	386	aValPheAlaGlyTrpThrAlaAlaLeuThrThrAspProAlaValLysLysAlaG	406
Db	1077	GTTTGGCTTGGGA---CGCGTGGCGACCCCGACTCTTCACCGCGGGCGCTGGGGCGGC	1133
QY	406	uSerPheIIeGlnAspThrValLysSerThrLAspSerThrLysSerThrIglTyValAlaAla	426
Db	1134	TGCGGTGTGGCCAAAGC-----GAGCGCGCTGTGGAGTGTGGCGGGGTGGCGG	1187
QY	426	pGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGlnAlaIleThrHISThr	446
Db	1188	GGCACAC-----CGCGCGGGGTCCGGCGGGGCGACACGGCGACATCTGGCGCCAG	1238
QY	446	rgLIyAlaSerLeuArg-----AsnThrValAsn	455

```

|||||:|||||
Db 1239 GGGAGCCGCGACTGGCGACTGTTCGACACCGGTCAAC 1272
|||||:|||||
RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: TUBERCULOSIS
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Alignment Scores:
Pred. No.: 247 Length: 4403765
Score: 132.50 Matches: 123
Percent Similarity: 34.23% Conservative: 55
Best Local Similarity: 23.65% Mismatches: 195
Query Match: 5.49% Indels: 147
DB: 4 Gaps: 21
US-09-825-414-7 (1-486) x US-09-103-840A-2 (1-4403765)
QY 20 ThrAlaSerASPAlaSerLeuAlaSerLeuAlaSerValSer----- 35
||||| ||| :|||||
Db 1632092 ACCGCGCTGCGCGCGCGCGCGCGGGGTTACCGCGCTTACCGCGCGACCGCGCGTTCACATC 1632151
|||||:|||||
QY 36 SerAspGInGlnArgGluGluLeuAlaLeuAlaLeuAlaSerTyrLeuThrAspHisValaPheAla 55
|||||:|||||
Db 1632152 GCGGAAGCGAAGTTCGCGCGTGGCGCGCGTTCGCGCGACCGCGCGAGCGCGCGCGCGCC 1632211
|||||:|||||
QY 56 AlaHisLysLeuProProAlaAspSerAlaAspGluGlnAlaValaAspValHisAsn 75
|||||:|||||
Db 1632212 CCTTTGCGCGCGGACCGCGCGCGACACCGCTGGATTCCGTTGGCGCAAGAGGTTCCCGCC 1632271
|||||:|||||
QY 76 -----AlaGlnLeuThrAlaLeuLeuGluThrArgAlaSerArg 88
|||||:|||||
Db 1632272 CAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632316
|||||:|||||
QY 89 LeuHisPheGluGluGluThrProAlaThrLeuAlaAspThrPheAlaLysAlaGluLys 108
|||||:|||||
Db 1632317 -----GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632364
|||||:|||||
QY 109 LeuAspArgLeu-----AlaThrThrThrSerGluAlaLeuArgAlaThrPro 124
|||||:|||||
Db 1632365 GTTGGCGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632424
|||||:|||||
QY 125 PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaLeuAlaAsnLysGluAspTrpLeu 144
|||||:|||||
Db 1632425 ACCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632463
|||||:|||||
QY 145 ProAlaProLeuLysProLeuThrProLeuHisSerGluAlaLeuSerGluAlaMetAsp 164
|||||:|||||
Db 1632464 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632523
|||||:|||||
QY 165 GlnValGluThrLysMetMetAspArgAlaThrGluLysAspLeuHisTyrTyrLeuSerAlaSer 184
|||||:|||||

```

```

Db 1632524 GGTGGTGG-----CGAGGAACGGCC----- 1632544
QY 185 ProAspArgLeuHisAspAlaMetAlaSerValLysArgHisSerProSerLeuAla 204
Db 1632544 ----- 1632544
QY 205 ArgGlnValLeuAspThrGlyAlaAlaValGlnThrTyrSerAlaArgAsnAlaValArg 224
Db 1632545 -----GACACCGGCGCATGGCTCCCGCCCTTGGCGCC----- 1632577
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGly----- 239
Db 1632578 -----GGCCCCGCGCATTACCAACAAGCCGCGCTTGGCGCCCTTGGCGCC 1632628
QY 240 -----AlaValAspLeuGlyValSer-----MetAlaGlyGlyLeuAlaAlaAsn 254
Db 1632629 GCCGCGCCCGCGCGCAGCGAGGCGTTCGGCGCGCTTGCAGCGGCGTGGCGCGCGTTCGCGC 1632688
QY 255 AlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAla 274
Db 1632689 GCTGTGTGGGGGTGGCGCGCGCAGCAGCATGCTGCACCCGCGGTGATGC--GGGCGCGCC 1632745
QY 275 LeuValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGlnGluAsnSptPrr 294
Db 1632746 GATCCACACACGACACCCAGGATGGCGCGGCTACCGCGCTTGCAGCGCGTTGCCACCAATCC 1632805
QY 295 LeuGlnAlaTyr-----LysAlaIleLysSerAlaSerTyrSer 307
Db 1632806 ATGCGCGCGCGCTTATCGAAGCGGCCCTTGGCAGCGTTCGCGCGCATCAGCCGCGCATGCGCGC 1632865
QY 308 GlyAlaAlaLeuAsnAlaGly----- 314
Db 1632866 GGCACCGCGCGTTCGCGCGCGCGCGCGCGCACCCATGCTGCGCTCGTGGTGGCGTGC 1632925
QY 315 LysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSer 334
Db 1632926 AAGGCGCTTACCGCGCTTGCACCGCGCTCCACCGCGCACCGCGCGCTCCACCGTTGCGCGC 1632985
QY 335 LeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGly 354
Db 1632986 CTTGCGCGCGCGTGGTGCATCCGCGCGCTGCGCCCGCAGGCGCTTAAGCGCGGTGGCGCGGT 1633045
QY 355 ValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAla 374
Db 1633046 -----GGCGCGCTTGGCGCGC 1633060
QY 375 ValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAla--GlyTyrPheThr 393
Db 1633061 GTTCCCGCGCTTACCGCGCGCGCGCGCGCACCGCGCTCGCTTGGCTCCGCGCGTC 1633120
QY 394 AlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPhe----- 408
Db 1633121 GCGCGCTTACCGCGCAGCGCGCGCGCACAGCTCCGCGCGCACACGCGCTTAAGGTGGCGCAGA 1633180
QY 409 IleGlnAspThrValLysSerThrAlaSerSerThrThrGly--TyrValAlaAspGln 427
Db 1633181 AGCGGCGGCGACGCGGCGCGCTTGGCGCGCGCACCTCCGCGCGCGCATTGGCGCAGAG 1633240
QY 428 ThrValLysLeuAlaLysThrValLysAspMetGlyGlyGlnAlaIleMetHisThrGly 447
Db 1633241 CCGCGCGCGCGCGC-----GGCGCGCGCGTTAC--GCCGCGC 1633276
QY 448 AlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProAla----- 462
Db 1633277 TTTGCCGCGCGCATGAGAAGTGGCGCGCTTCCCGCGCGCGCTTGGCGCGCGGTGT 1633336
QY 463 -----ArgGlnAlaAspIleGlnGlyGlyThrAlaAlaSerProSer 477
Db 1633337 GGGCGTGGCGCCGCGCGCGCGTGGCGCACGATGCTGGAGCCGCGCTCCGCGGTGCTCC 1633396

```

```

: Patent No.5876991
:
: GENERAL INFORMATION:
: APPLICANT: Dehoff, Bradley S.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rostock, Paul R., Jr.
: APPLICANT: Sutton, Kimberly L.
: TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
: NUMBER OF SEQUENCES: 15
:
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: THOMAS G. PLANT 1501
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: ASCII(DOS) text only
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,227C
: FILING DATE: February 21, 1997
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Plant, Thomas, G.
: REGISTRATION NUMBER: 35,784
: REFERENCE/DOCKET NUMBER: X-8231
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-2459
:
: INFORMATION FOR SEQ ID NO: 7:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 44377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..14002
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14046..20036
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20110..31284
:
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31329..36071
:
: NAME/KEY: CDS
: LOCATION: 36155..41830
:
: US-08-804-227C-7
:
: Alignment Scores:
: Pred. No.: 0.228 Length: 44377
: Score: 130.50 Matches: 135
: Percent Similarity: 33.77% Conservative: 73
: Best Local Similarity: 21.92% Mismatches: 214
: Query Match: 5.41% Indels: 195
: DB: Gaps: 28
:
: US-09-825-414-7 (1-486) x US-08-804-227C-7 (1-44377)
:
: Oy 12 ValTrrAlaThrAspSerPhearghhrAlaSerAspAlaSerAspAlaSerSerVal 31
: ::::||||| :::: ||||||||||||||| :::: ||| ::::
: Db 18069 CTGACCCACCAACGCGACGCGCGCTACGGCGTGGACGGGGG---CGGCTCAGTCTCTC 18125
: :
: Oy 32 ArgSerValSerSerAspPcInArgJuIlleAsnAlaIleAlaAspTyrLeuThrAsp 51
: :||| :::: ||| :::: ||| :::: ||| :::: ||| :::: |||
: Db 18125 GACGAGGTCGTCGACGAGGTGCGCGGACCCCTCGAAGCCGTACAGCCCGCTCGCGCAG 18185
: :
: Oy 52 HisVal-----PheAlaIaHis 57

```

Db	18186	ACCGAAACGGGCCCCCGAGCTGCAGCTCCGTAACGGCGGGGCGGCCCCCGGACAGACGGCCGT	18245
Qy	58	LysLeuPProPProAlaAspSerAlaAspIyGlnAlaValAspValHisAsnAlaGln	77
Db	18246	ACAAACCCCCCGGTGGACACCGCGACGGAGCCCGACCGCGTAC-----GGCCCCGG	18299
Qy	78	IleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGly-----GluThr	95
Db	18300	CTTCGTCTCTGACCCCGGGGCGGCGCGGACCC-----GAGGAGGCGCGGCGCAT	18350
Qy	96	ProAlaThrIleAla-----AspThrPheAlaIyAspAlaIyLysLeuAspArg	111
Db	18351	CCCCGGGGTCCGCTGTCTGGGGGCTCGTCCGGGTCGCCCAAGCCGCAAGCCGCGCCG	18410
Qy	112	LeuAlaThrThr-----ThrSerGlyAlaLeuArgAlaThrProPheAlaMet	127
Db	18411	TTACCTCTGTGTGAGCTGCAGCGCACCCAGCGCTGTGTGGGCGCTCGCTCGTGG	18470
Qy	128	AlaSerLeuLeuGlnIyTyrMetGlnProAlaIleAsnLysGlyAspThrLeuProAlaPro	147
Db	18471	GGCAGG-----GATCGCGGCAATGTGGCCGTGGCGGACGAGTGTACCGTCCCGGC	18524
Qy	148	LeuLysProLeuThrProLeuIle-----	155
Db	18525	CTCGTCCGGGTGGCGGACCGGCTCCCGCACAGCGGGGCGACCGCGCGGACGCGGCT	18584
Qy	156	SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThr	175
Db	18585	GGCGGCGAGCGGCTCGGACCGTGGACCCCGGACCGGACCGCTGTATACCGCGGCGGACC	18644
Qy	176	GlyAspLeuHisTyrLeuSerLysSerProAspArgLeuHisAspAlaMetAlaLysSer	195
Db	18645	GGA-----GCACGTGGCGCGGAA	18662
Qy	196	ValLysArgHis-----	199
Db	18663	ACCGCGCGGACCTGTGCAGCCGCGACAAAGTGGCCCATCTCTCTGTGTGGGACGGCC	18722
Qy	200	-----SerProSerLeuAlaArgGlnValLeuAsp-----ThrGly	211
Db	18723	GGTCCCGACGCAACCGCGGTGATGATGACTGTGCGCGGATGTACCGAGTGGGGTCCGAG	18782
Qy	212	ValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArgThrValLeuAlaPro	229
Db	18783	GTCGCGTGAAGGGCTGTGACGTGCAGCGACCGGCGCGCTCGCGGCTGTGCAGCA	18842
Qy	230	AlaLeuAlaSerArgProAla-----ValGlnGlyAlaValAspLeuGly	244
Db	18843	CTCCCCGACGAACCCCGCTGACCTGCGTGTGGCACACCGCGGGGTGTGTGCACAGCGC	18902
Qy	245	ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal	264
Db	18903	GTC-----CTCTCCGCC	18914
Qy	265	GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro	284
Db	18915	CAGAGCGCGGACGG--GATGTGACACGCTCTCCGCGCCAAAGCGGACGCGCGCTCACCT	18973
Qy	285	---LysAlaGlnLeuSerGluGlnAsnAspThr--LeuGlnAlaTyrLysAlaIleLysLeu	303
Db	18974	GGACGAGCTGACCCGGGAGATCGGACGGGCTGCCCTGTGTACTCTCGTCTCGTCCGC	19033
Qy	303	r-----AlaSerTyrSerGlyAla-----AlaLe	311
Db	19034	CACCTTGGGCAACGGGGGCGAGCGGGGTACCGCGCGGCCCAACGCTTCATGGACCGCT	19093
Qy	311	uAsnAlaGlyLysArgMetAlaGlyLeuPro-----	321
Db	19094	GAGCCCGCGGCGGTGCGCGCGGCGGACCCCGCGCTGTGCGTCTGGGCTGGGTGGTGC	19153
Qy	322	---LeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSe	340

```

Db 19154 CGGCGTGGCTCTCGCACCGGACTGAGAGAGCG-----GACGCGGCGG 19198
Oy 340 rLeuthrGlnasnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnI 360
Db 19199 GGTTCAGCGCTCGGGCTCTCCGCGCCGCTCGACCGCGCGCCGACCTCGACCG 19258
Oy 360 uMetAlaThrLysAsnIleThrAspProAlaThrLysAlaIaValSerGlnLeuThrAs 380
Db 19259 GCGCGTGAACCGG-----CCGACCGCGGCTCTGTCGCCGTGCGGCTCGACCTGGCGG 19312
Oy 380 nLeuAlaGlySerAlaIaVal-----PheAlaGlyIlePThrThrAl 394
Db 19313 CGCGCGCGGTGCCACCGCTCTCCGAGGTCTCTGCTGATCTGCGCGCGGTACCGGCGGA 19372
Oy 394 aAlaLeuThrTh-----AspProAlaValLysIy 404
Db 19373 CGCGCGGACGACCGCGGCGCGCGGCGGACCGAGGACGAGACGCTGCGTGGCGCC 19432
Oy 404 sala----- 405
Db 19433 TGCACCGCGCGCGCGGACCGCGGACGCTGGCGCGGCTCGCGGAGCGTTCGG 19492
Oy 406 -----GluSerPheIleGlnAspThrValLysSerThrAlaSerThrIle 422
Db 19493 ACCCGACCGTACGGCTCTCTCTCTGACCTGTGTGGACCGAGGTCCGCGGCGGTCTCGG 19552
Oy 422 yTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAla 442
Db 19553 ACACGGCGACCGCGCGCGGATCGGCGCGCGCGCACCTTCMAAGGACCGCGATTTCGACTC 19612
Oy 442 aIleThrHisThrGlyAlaSerLeuArgAsnThrAlaSn-----AsnLeuArg-- 458
Db 19613 CCTC-----ACCGCTGTGACCTCGCAACCGGCTGAACACGACACCGGACTGCGGCT 19666
Oy 459 -----GlnArgProAlaArgLysAlaAsp----- 466
Db 19667 GCCCGGACCGCTGTGTGGACACCGCACACCGCTGCCCTCGCGCACTCTGCTCGCA 19726
Oy 467 ----IleGlnGlyGlyThrAlaAlaSerProSerGlnIlePro 480
Db 19727 CGGCGTGAAGCGCGGCTCCAGCGGAACCGGCTGATGAGTCCGG 19772

RESULT 11
US-08-804-198-1
: Sequence 1, Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Kuhstoss, Stuart A.
: APPLICANT: Rao, Nagendra R.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Rostock, Paul R., JR.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,198
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CANTRELL, PAUL R.
: REGISTRATION NUMBER: 36,470

```

Db 19727 CGGGCTGAGGCGGCCGCTCAGCGAACC GGCGCGTGTAGGTCCG 19772

RESULT 11
US-08-804-198-1
Sequence 1, Application US/08804198
Patent No. 5945320

GENERAL INFORMATION:
APPLICANT: Burgett, Stanley G.
APPLICANT: Kustoss, Stuart A.
APPLICANT: Rao, Nagaraja R.
APPLICANT: Richardson, Mark A.
APPLICANT: Rosteck, Paul R., Jr.
TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: PAUL R. CANTRELL 1138
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: Macintosh 7.0
SOFTWARE: Microsoft Word 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,198
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CANTRELL, PAUL R.
REGISTRATION NUMBER: 36,470


```

Db 19553 ACACGGCAGCCCGGCGATCGGCGGCCGCCGCACCTTCAAGGACGCCGATTCACATC 19612
Oy 442 attleThHsthrlyAlaserlAserArgAsnrThrValAsn-----AsnLeuArg-- 458
:::||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 19613 CCTC-----ACCCTGTGCACTTCGCCAACCGGCTGAACAACAAGCACCGACTGGGGCT 19666
Oy 459 -----GlnArgProAlaArgGluAlaAsp----- 466
Db 19667 GCCCGCAGCCCTGCTTCGACACACCCACACCGCTTCGCCCTGCAACTCCTGCTCGA 19726
Oy 467 ----llglugluglygylThrAlaAlaserProSerGluIlePro 480
Db 19727 CGGGCTGSAGCGCGCGCTCCAGCGAAACGGCCCGCTGAGGTCCCG 19772

RESULT 12
US-08-961-527-363
; Sequence 363, Application US/08961527
; Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 Inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 363:
SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-363

Alignment Scores:
Pred. No.: 0.00788 Length: 4483
Score: 129.00 Matches: 96
Percent Similarity: 35.98% Conservatve: 81
Best Local Similarity: 19.51% Mismatches: 269
Query Match: 5.35% Indels: 48
DB: 4 Gaps: 8

US-09-825-414-7 (1-486) x US-08-961-527-363 (1-4483)
Oy 5 ArgArYValGInGInProValThrAlaThAspSerPheArGThrAlaSerAspAla 24
|||:|||:|||:||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1382 CGACAGAGTGCGTGGCTCA-AACAGTGCATCTGATCGGCATCAACACAGCTGGTCAGCC 1440
25 SerLeuAlaserSerValArgSerValSerSeraSpclnGlnArgGluIleAla 44
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

[illegible]

Qy	386	avAlPheAlaGLytrPphrThrAlaAlaLeuThrThrAspProAlaValLysAlaGl	406
		
Db	2502	TACCAAGCGCTCA---GCCACGACCAAGCACACAGTCCAGCTTCAGCAAGTACCAAGTGC	2558
Qy	406	userPheIlleolnAspThrValLysSerThrAlaSerSerThrPheGlyTrValAlaAs	426
		
Db	2559	GTCAGCCTCAGCGTCGACCAAGTCCGTCGGGTTCCAGCAATACCTCAGGTCGAAATCAGC	2618
Qy	426	pGlntHrValLysLeuAlaLysThrValLysAspMetGlyGluAlaIleThrHisTh	446
		
Db	2619	ATCAACAGATGCATCAGCTTCAGCAATCAACAAAGTCCAGCTTCAGCAAGTACCAAGTGC	2678
Qy	446	rGlYAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGluAlaAs	466
		
Db	2679	GTCGGCTTCAGACCAACAGAGTGCTTCAGTCCAGGCTCAACCAAGTCCCTCGAATCAGC	2738
Qy	466	pIleGluGluGlyLysThrAlaAlaSerProSer	477
		
Db	2739	ATCAACAAGTCCCTCGGCTTCAGCAAGCAACCACT	2772

```

RESULT 13
US-09-320-878-19/C
: Sequence 19, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 300622002120
: CURRENT APPLICATION NUMBER: US/09/320,878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141,908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073,538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846,247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119,139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100,880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 19
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

Alignment Scores:		
Pred. No.:	0.29	length: 38506
Score:	128.50	Matches: 115
Percent Similarity:	36.31%	Conservative: 64
Best Local Similarity:	23.33%	Mismatches: 195
Query Matchn:	5.33%	Indels: 123
Delet:	3	Gaps: 19

QY	4	ASfATgATgValGInGInProPro---	ValMhralatHnAspSerPheargHhAlaser	22
Db	22213	TGCGGACATGCTCGCGCCACCGGGGGCGCGGGGCAAGCCACCAAGAGCGTGGC		22154
QY	23	ASpAlaserLeu-----AlaserSerValArgSerValSerSerApGInGIn	39	
Db	22153	GCGGCGGGGGTTCGCCCCCGGTGCAGCGGGCTTCACACAGTCCGTAAAGTTCGGGAAGGAC	22094	

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:32:35 ; Search time 76.9209 Seconds
(without alignments)
2836.537 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHNRVOOPVTATDSFRT.....IEGGTAASPEIIPRPMRS 486

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 396772 seqs, 224632407 residues
Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=rlp
-Q=/cgn2.1/USPTO.spool/US09825414/rnat.28012003.161742.26139/app.query.fasta.1.1294
-DB=published_Applications_NA -QFWT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plc -NORM=ext -HEADSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09825414.@CGN.1.1.46 @rnat.28012003.161742.26139
-NCPUP=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published_Applications_NA:*

1: /cgn2.6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2.6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2.6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2.6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2.6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2.6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2.6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2.6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2.6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2.6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2.6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2.6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2.6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2.6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2412	100.0	1461	10	US-09-825-414-6
2	2412	100.0	30365	10	US-09-825-414-1
3	1869.5	77.5	1464	10	US-09-825-414-65
4	148.5	6.2	4645	9	US-09-971-536-27

5	143	5.9	1801	10	US-09-287-849-21	Sequence 21, Appl
6	133.5	5.5	2287	10	US-09-287-849-1	Sequence 1, Appl
7	133.5	5.5	5036	10	US-09-918-951-2	Sequence 2, Appl
8	131.5	5.5	6258	10	US-09-815-242-4584	Sequence 4584, Ap
9	131.5	5.5	17388	10	US-09-815-242-8512	Sequence 8512, Ap
10	129.5	5.4	7035	10	US-09-815-242-8615	Sequence 8615, Ap
11	129	5.3	7302	10	US-09-815-242-4780	Sequence 4780, Ap
12	129	5.3	18846	10	US-09-815-242-4898	Sequence 8898, Ap
13	127.5	5.3	11797	10	US-09-287-849-25	Sequence 25, Appl
14	127.5	5.3	11220	9	US-09-860-846-32	Sequence 32, Appl
15	127.5	5.3	11220	10	US-09-861-289-32	Sequence 32, Appl
16	127.5	5.3	15872	9	US-09-860-846-1	Sequence 1, Appl
17	127.5	5.3	15872	10	US-09-860-846-1	Sequence 1, Appl
18	127.5	5.3	36778	9	US-09-860-846-5	Sequence 5, Appl
19	127.5	5.3	36778	10	US-09-861-289-5	Sequence 5, Appl
20	127	5.3	1248	10	US-09-974-300-902	Sequence 902, App
21	127	5.3	3300	10	US-09-379-931-6	Sequence 6, Appl
22	127	5.3	9542	10	US-09-813-214A-8	Sequence 8, Appl
23	126.5	5.2	32768	10	US-09-070-927A-399	Sequence 399, App
24	125	5.2	2141	10	US-09-841-786-11	Sequence 11, Appl
25	125	5.2	9726	10	US-09-841-786-15	Sequence 8, Appl
26	125	5.2	11130	10	US-09-841-786-15	Sequence 15, Appl
27	124.5	5.2	7995	10	US-09-864-864-333	Sequence 333, App
28	123.5	5.1	6177	9	US-10-124-800-3	Sequence 3, Appl
29	123.5	5.1	8078	10	US-09-791-942-3	Sequence 3, Appl
30	120.5	5.0	7639	9	US-09-971-536-28	Sequence 28, Appl
31	120	5.0	8730	9	US-10-124-800-1	Sequence 1, Appl
32	119.5	5.0	2155	10	US-09-960-428-13	Sequence 13, Appl
33	119.5	5.0	4645	9	US-09-971-536-27	Sequence 27, Appl
34	118.5	4.9	3954	10	US-09-815-242-4063	Sequence 4063, Ap
35	118	4.9	7100	10	US-09-932-183A-1	Sequence 1, Appl
36	117.5	4.9	2331	10	US-09-815-242-4037	Sequence 4037, Ap
37	117	4.9	9319	10	US-09-956-004-85	Sequence 85, Appl
38	116.5	4.8	1380	10	US-09-815-242-7856	Sequence 7856, Ap
39	116.5	4.8	9210	9	US-09-712-363-100	Sequence 100, App
40	116	4.8	2523	10	US-09-476-242-15	Sequence 15, Appl
41	116	4.8	7434	10	US-09-815-242-4761	Sequence 4761, Ap
42	116	4.8	7437	10	US-09-815-242-8869	Sequence 8869, Ap
43	116	4.8	8979	9	US-09-738-626-2739	Sequence 2739, Ap
44	116	4.8	68750	9	US-10-014-717-1	Sequence 1, Appl
45	115	4.8	1029	10	US-09-815-242-6946	Sequence 6946, Ap

ALIGNMENTS

RESULT 1
US-09-825-414-6
; Sequence 6, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collmer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-09-825-414-6

Alignment Scores:

Pred. No.: 1.03e-224 Length: 1461
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-825-414-7 (1-486) x US-09-825-414-6 (1-1461)

OY 1 MethiIAsnArgArgValGlnGlnProValThrAlaThrAspSerPheArgThr 20
DB 1 ATGCACATACACCGACCGCTCCACACACCGCTGTGACTGCGAGGATGACTTTCGACA 60
OY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerSpGlnArg 40
DB 61 GGGTCGACGCGCTCTTCCCTCCACCTGCTGCGATGTGTGCTCCGATCCGACACCG 120
OY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaIleHisLysLeuPro 60
DB 121 GAGATAAATGCGATGCCATTTACCTGACAGATCATGTGTCGCTCCGCTAAACTGCCG 180
OY 61 ProAlaAspSerAlaSpGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
DB 181 CCGGCCGATTCGGCTGATGCGCAAGCTGCACTGACATACATCGACATCGACTCGCG 240
OY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
DB 241 CTGATCGACAGCGCGCCAGCGCGCTGCACTTGGAAGGGGAACCCCGGACCATCGCG 300
OY 101 AspThrPheAlaAlaLysAlaLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
DB 301 GACACCTTCGCCAAGCGGAAAAAGCTGCACCGATTGGCGACGACTATCAGCGCGCTTG 360
OY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140
DB 361 CCGGCCACCGCCCTTTCCTCCATGGCTCTGTTCTTCAATGACGCTCGCATCAACAAG 420
OY 141 GlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
DB 421 GCGGATTTGGCTGCGCGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGAGCGCTG 480
OY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrLysPheLysIleTyr 180
DB 481 GGGCGCATGACCGATGGGCGACCAAGATGATGGACCGCGCGGATATCGCATCTGAC 540
OY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
DB 541 CTGACGCGCTCCGCGACAGCTCCACGATGCGATGCGCTTCGGTGAAGCCCACTCG 600
OY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
DB 601 CCAAGCGCTTCGTCGACAGCTTCGACACGGGGTTCGGTTCAGACGATCTGCGCGCG 660
OY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
DB 661 AACGCCGTACGACCGCTATTTGGCTCCGCGACGCGCTCCAGACCCCGCTGCAAGGTGCT 720
OY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
DB 721 GTGGACCTTGTTGATGATGAGTGGGGGTGTGTCGCTCCACCGACGCTTTGGCAACCG 780
OY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
DB 781 CTGCTCAGTGTGACGTGCGGTGATCACACGCGGGGTGATTAAGTTCGCTTGAAG 840
OY 281 AspLysGluProLysAlaGlnLeuSerGluGluAsnAspTyrLeuLysIleTyrLysAla 300
DB 841 GATTAAGACCCCAAGGCTCAACTGACGGAAGAAACGACTGTGCTCAAGGCTTATTAAGCA 900
OY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB 901 ATCAATATGCGCAGCTACTCGGGTGGCGCTCAACGCTGGCAAGCGGATGGCGGCTCG 960

OY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
DB 961 CCACGTGATATGGCGACCCACCAATGGGTGGGTAAAGCCCTGCTCACCGTCCACG 1020
OY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGly 360
DB 1021 CTGACCCAAAGCGTCTGCGCTCGCGGGGTGCTTTGGCAGGGGTAGCAAGTTGACAGAG 1080
OY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
DB 1081 ATGGCAGCAAAAATATTCACCGACCGCGGACCAAGCCGCGGTGATGATGACCAAC 1140
OY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrPThrThrAlaAlaLeuThrThrAspPro 400
DB 1141 CTGGCAGGTTTCCGACCGCTTTTCCAGAGCTGAGACACGCGCGCTGCAACCGATCC 1200
OY 401 AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
DB 1201 GCGGTGAAAAAGCCAGTCTTCATACAGGACACCGTGAATTCGACTCATCCACTAC 1260
OY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
DB 1261 ACAGGCTACGTACCGACCAAGCGCTCAACTGCGCAAGACCTTCAAAGACATGGCGGG 1320
OY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
DB 1321 GAGCGCATCACCATACCGGCGCGCTGCGCAATACGTCATTAACCTCGCGTCAACGC 1380
OY 461 ProAlaArgGluAlaAspIleGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
DB 1381 CCGGCTCGTGAAGCTGATATAGAAAGGGGGGCGACGCGGCTTCTCAAGTAAATACCG 1440
OY 481 PheArgProMetArgSer 486
DB 1441 TTTGGCTATGCGGTGCG 1458

RESULT 2

US-09-825-414-1/c

; Sequence 1, Application US/09825414

; Patent No. US20020083489A1

; GENERAL INFORMATION:

; APPLICANT: Collier, Alan

; APPLICANT: Alfano, James R.

; APPLICANT: Charkowski, Amy O.

; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE

; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES

; FILE REFERENCE: 19603/3243

; CURRENT APPLICATION NUMBER: US/09/825,414

; PRIOR FILING DATE: 2001-04-03

; PRIOR APPLICATION NUMBER: 60/194,160

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: 60/224,604

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: 60/249,548

; PRIOR FILING DATE: 2000-11-17

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 30365

; TYPE: DNA

; ORGANISM: Pseudomonas syringae

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (29734)

; OTHER INFORMATION: n at any position is undefined

US-09-825-414-1

Alignment Scores:
Pred. No.: 6.05e-223 Length: 30365
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0


```

Db 61 GCTTCGAGCGCTCCCTGCTGAGTCTGTGCGGCTGTGACACTACCTGTCGCGC 120
QY 41 GIU11LeuAla1LeuAlaSerYrLeuThrAspHisValPheAlaIleHisLysLeuPro 60
Db 121 GATCTACAAAGCTATTCCGCTATTCTGAAACATCACTGCTTCCTGCGCACAGCTTTTCG 180
QY 61 ProAlaSerSerAlaSpGlnAlaValAlaValAspValHisAsnAlaGlnIleThrAla 80
Db 181 GTAATAGGCTCACCGGATGAGCTGATGCGGCTTCGACACAAAGAGAGATCGATGCGC 240
QY 81 Leu11LeuThrAlaSerArgLeuHisPheGlnLys11LeuThrProAlaThr11LeuAla 100
Db 241 TTGGTAGACAGACGCGCAACCGCTGTACTCCGAAAGGAGAACCCCGCAACATGCGCC 300
QY 101 AspThrPheAlaLysAlaGlnLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu 120
Db 301 GAACATTTCCGCAAGGCGGAAAGTTGCGAGCTTTGGCGAGACCGGATCAAGTCTTTT 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuGlnIleThrMetGlnProAla11LeuAla 140
Db 361 GAGAACACGCGCATTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
QY 141 GlyAspThrLeuProAlaProLeuLysProLeuThrProLeu11SerGlyAlaLeuSer 160
Db 421 GCGGATTTGCTGAGCAACGCGCTCAAGCCGCTGACCCCTCATTTTCCGAGCGCTGTCG 480
QY 161 G11ValaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisLys 180
Db 481 GGAACCATTCAGACGAGTGGCGACCAAAATAGATGATGCGAGGGGTGATTTGCAATTAC 540
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
Db 541 CTGAGCACTTCGCGGAGCATTTGCATGATGCGATGCGCCGATGCGGAAAGCGGCACTCG 600
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnIleThrYrSerAlaArg 220
Db 601 CCTCGCTTGTGTCACAGGTGGTGGACATGCGGATTCAGTGCAGACGTTCTCGGCGCTA 660
QY 221 AsnAlaValAlaArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 661 AATGTGCTGCGTACCGATTTGCTCCGACACTACGCTCCAGACCGTGGTGGAGGCT 720
QY 241 ValaSpLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 721 GTTCATTTTGGCGTATCTACGCGGCTGGCTGGTGGTGGCAATGCAAGCTTTGGCAGCGC 780
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLysLys 280
Db 781 ATGCTCAGTGTGCAATGCGCGATCAACTGCGTGGGGGCGATTGTAATTGGCATGAAA 840
QY 281 AspLysGlnProLysAlaGlnLeuSerGlnLysAsnAspTrpLeuGlnAlaTyrLysAla 300
Db 841 GATTAAGAGCCCAAGCGCGCTTACGTAGAAACACTGATTCGCTTGAATGCTTCAAAAGCG 900
QY 301 IleLysSerAlaSerYrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 901 ATCAAGTCGCGCACTACTAGTGTGCGGCTCAATGCGGCGCAAGCGGATGCGCGGCTG 960
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 961 CCATGTGACGTCGCGGACGCGGCTCAAGGCGGTGAGAACTGTGTGTGCGGCACACAGC 1020
QY 341 LeuThrGlnAsnGlnLysLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnLys 360
Db 1021 CTGCAAAAATGCGCTGCGCTGAGCCGCTGAGCGGTGTTACCGCGGGGTAAAGTTCACAGAAA 1080
QY 361 MetaLathrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
Db 1081 ATGCGCAGAAATAATACACTGATTTCGCGCACCAAGGCTGGTGTAGTCACTGAGCAAC 1140
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrThrAspPro 400

```

```

Db 1141 CTGTGGGTGCTGAGCGCTTTTCGAGGCTGAGCAACCGGCTGAGTGGCTGACCTT 1200
QY 401 AlaValLysLysAlaGlnSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 1201 GCGGTAAAGAACCCAGATCGTTTATACAGATTAAGTGAATCGAACCCGATTTATAC 1260
QY 421 ThrGlyTrpValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 1261 ACAAGCATATGTCGCGCACGACCGCTCAAACTGCGAAAAACATGCAAGGACATGACCGG 1320
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 1321 GAGCGCTTCACGACCGGCTGCGACCTTACGCGAGTCTCATTAACCTGCTCATTCG 1380
QY 461 ProAlaArgGlnAlaAspIleGlnGlyGlyThrAlaAla---SerProSerGlnIle 479
Db 1381 TTGCGTCCGGAAGCTGATATACGAAGAAAGTGGGATTTGCGCGCTTTCTCGAATGAAACA 1440
QY 480 ProPheArgProMetArg 485
Db 1441 CCGTTTCAGCTCAGCGCT 1458

RESULT 4
US-09-971-536-27
; Sequence 27, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 4645
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus

US-09-971-536-27

Alignment Scores:
Pred. No.: 7,58e-05 Length: 4645
Score: 148.50 Matches: 126
Percent Similarity: 36.18% Conservative: 69
Best Local Similarity: 23.38% Mismatches: 233
Query Match: 6.16% Indels: 112
DB: 9 Gaps: 21

US-09-825-414-7 (1-486) x US-09-971-536-27 (1-4645)

QY 13 ThrAlaThrAspSerPheArgThrAla-----SerAspAlaSerLeuAlaSerSer--- 29
Db 2864 ACTGCAACATCTCGTATGTGTAGTGTGCAAGTTCGAGTTCGCACTGAGTGAAGCGGTCTATAT 2923
QY 30 -----SerValArgSerValSerSerAspGlnArgGlnIleAsnAlaIleAlaAsp 47
Db 2924 GCTCATGAGCGAGGCTCCAGCGCGATGACGCTGTGCGTCACTCGGCGAGTGCAGCCAA 2983

```

```

QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProPro-----AlaAspSer 64
   :::: :::: ||||| ||||| |||||
Db 2984 CATGCCAGCAGCCGCTGCAGTCCGCGCATCCAGTATCCGAAGATAGTGGATTCAGTCA 3043
QY 65 AlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGlyThr 84
   ||||| ||||| :::: :::: ||||| |||||
Db 3044 CTAGCCATGCTAGCTGCAGCAGCGCAGCAAGCAAGCAAGTAAAGCAAGTATGCCGCGCAC 3103
QY 85 ArgAlaSerArgLeuHisPheGlnGlyGlyThrProAlaThr-----IleAlaAspThrPhe 103
   ||||| :::: ||||| :::: ||||| :::: |||||
Db 3104 AGCGCCGCGCGCTGGTGTTCAGTCTGCATGCTGCCAGTATGCAAGTAAACAGCGCAAGCGGCT 3163
QY 104 AlaAlaGlnAlaGlyLysLeuAspArgLeuAlaThrThr-----SerGly 118
   ||||| :::: ||||| :::: ||||| :::: |||||
Db 3164 GCAGAGCGCGATGTGGTGGCAAGCAAGTGGCGCCAGCAGCGGTAACTAACGTAATGCCAGTGGC 3223
QY 119 AlaLeuAlaArgAlaThrPro-----PheAlaMetAlaSer 129
   ||||| ||||| ||||| ||||| |||||
Db 3224 GCAGCCAGTGGCAGCAAGGCTGTGATAGCAAGCCGCAAGCAGAGATTTCTGATGTCAGCG 3283
QY 130 -LeuLeuGlnIleGlyMetGlnProAlaIleAsnLysGlyAspThrProAla----- 146
   :::: ||||| ||||| :::: ||||| ||||| |||||
Db 3284 AGTGTCTGACAGCAAGCAAGTGTCCAGCGTGCAGAAAG-----CAATTGCCAGCAAGCGCGC 3337
QY 147 ----ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla-----Me 163
   ||||| ||||| ||||| ||||| ||||| |||||
Db 3338 AGTGGCGCGCGCATCCGATGATCTCG-GTAGCTTTAGTCCGCGCAGTGGCGCTCAGCGCTT 3396
QY 163 LAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisIleLysSerAl 183
   ||||| :::: ||||| ||||| ||||| |||||
Db 3397 TGACAAAGCTGCCAGCGCTGGCGAAGCGCGCATCTCA-----AGTGC 3438
QY 183 AspProAlaArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerIe 203
   ||||| ||||| ||||| ||||| ||||| |||||
Db 3439 CCGGACCGCGCTGTAGTTCACCGCGCATCAGGCAAGCAGAGTGGCGCAAGTCCAG 3498
QY 203 uAlaArgGlnValLeuAspThrLysValAlaValGlnThr----- 217
   ||||| ||||| ||||| ||||| |||||
Db 3499 TGGCAGCGAAGCGGTCAAGCATCAACCGCACATCTGTGTATGCTATGCTGCAGAGTTTC 3558
QY 218 -----SerAlaArgAsnAlaVal 223
   ||||| ||||| ||||| ||||| |||||
Db 3559 GAGTGCAGTGAAGCGCGTTCATATGCTCATGCGCAGCGCTCCAGTCCAGTAAACGAC 3618
QY 223 L-----ArgThrValLeuAlaProAlaLeuAlaSerAlrPr 235
   ||||| ||||| ||||| ||||| ||||| |||||
Db 3619 TGGCCATGCAAGTATGCTACAAAGTCAAGCAAGTCCGATCCAGTGTGCTGCCAGGTA 3678
QY 235 oAlaValGlnGlyAlaValAspLeuLysLysMetAlaGlnGlyLysAlaAlaAlaAla 255
   :::: ||||| ||||| ||||| ||||| |||||
Db 3679 CCCAACTGATAGTGGATCCAGTCAAGTATATGCTGCCAGTGCAGCAAGTACTCTGC 3738
QY 255 agLysPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyLysAla 275
   ||||| ||||| ||||| ||||| ||||| |||||
Db 3739 ATCC-----AGTCCCGCTACTGCCCGCAGCAAAAGTGAAGCTTCGACGGC 3780
QY 275 uValLeuGlyLeuLysAspLysGlnProLysAlaGlnLeuSerGlnLysAsnAspThrPle 295
   :::: ||||| ||||| ||||| ||||| |||||
Db 3781 ATCGTCGGCTGCAGATCATGCTAGTGAACAAGCAAGTATTTGCTTCCAGTGAAGATGTGG 3840
QY 295 uGlnAlaArgLysAlaAlaLysSerAlaSerIleSerGlyAlaAlaLysAsnAlaLys 315
   :::: ||||| ||||| ||||| ||||| ||||| |||||
Db 3841 ATCAACAGTGTGCTGCGAGTGTCCGCTGCAGCGCGCGCGCGTCCAGTCCAGTCTCAAA 3900
QY 315 s-----ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAl 331
   ||||| ||||| ||||| ||||| ||||| |||||
Db 3901 GGCTGTAAACAGTATGCTGCGGTATATCTCATTCAGCAAGTGCAGAGCTCCAGACAG 3960
QY 331 aValAlaGSerLeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly 351
   :::: ||||| ||||| ||||| ||||| ||||| |||||
Db 3961 TGTAGAGCGCTGAAGTCAAGCAAGCAGTGTCCCGCGCTGCTGCT----- 4009

```

```

QY 351 yPheAlaGlyValGlyLysLeuGlnLysMetAlaThrLysAsnIleThrAspProAlaThr 371
Db 4010 -----TGTGATGATGCGGTACTTCAAGCTTCG 4038
QY 371 rLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyThr 391
   ||||| ||||| :::: ||||| ||||| |||||
Db 4039 CAGTGGCGCTGTTGCTGACGATGCTAAAGCAAGTACCGCGCGAGTATGATCCAGCGC 4098
QY 391 pThrThrAlaAlaLeuThrThrAspProAlaVal-----LysIle 404
   ||||| ||||| ||||| ||||| ||||| |||||
Db 4099 TACAACTGCTGCATAGTATGCTCCACATCTCTGGCTGATCAGATGCCAGCGTCAAC 4158
QY 404 sAlaGlySerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyThrVal 424
   ||||| ||||| ||||| ||||| ||||| |||||
Db 4159 CGCTGCTCCCATAT-----TTGCCAACTACTGTGGAGAAACAGCAGAGTATGAT 4209
QY 424 L--AlaAspGlnThr-----ValLysLeuAlaLysThrVal 435
   :::: ||||| ||||| ||||| ||||| |||||
Db 4210 ACCATGGGTGCAGACCGCAACACAGACAGCAAGCCACAGACAGCAAGCAACAAAGCA 4269
QY 435 LysAspMetGlyGlyGlnAlaIleThrHisThrGlyAlaSerLeuAlaArgAsnThrValAs 455
   ||||| ||||| ||||| ||||| ||||| |||||
Db 4270 GCCAACACAAAGCCGCTCAACA-----ACCGAGACAGGT--TCATTTACCGCAAGCGATCA 4323
QY 455 nAsnLeuArgGlnArgProAlaArgGlyLysAlaAspIleGlnGlyGlyThrAla 473
Db 4324 TGCAGGAGAGCATATGCTTACCGCAGACCGGTGATGCTGAAGAGGCTACTTCT 4378

RESULT 5
US-09-287-849-21
; Sequence 21, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287, 849
; PRIORITY FILING DATE: 1999-04-07
; PRIORITY FILING DATE: US 08/818, 112
; PRIORITY FILING DATE: 1997-03-13
; PRIORITY FILING DATE: US 08/942,578
; PRIORITY FILING DATE: 1997-10-01
; PRIORITY FILING DATE: US 09/025,197
; PRIORITY FILING DATE: 1998-02-18
; PRIORITY APPLICATION NUMBER: US 09/056,556
; PRIORITY FILING DATE: 1998-04-07
; PRIORITY APPLICATION NUMBER: US 09/223,040
; PRIORITY FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: cti-fusion
; NAME/KEY: CDS
; LOCATION: (1)..(1800)
US-09-287-849-21

Alignment Scores:
Pred. No.: 7,24e-05
Score: 143.00
Percent Similarity: 34.45%
Best Local Similarity: 23.03%
Query Match: 5.93%
Length: 1801
Matches: 117
Conservative: 58
Mismatch: 165
Indels: 168

```

DB:	10	Gaps:	25
US-09-825-414-7 (1-486) x US-09-287-849-21 (1-1801)			
Oy	20	ThrlaseraspalaserleualaserseerValargserValaserplngln	39
		:::::	
Db	262	ACCGGGGGGAGGCGGAGCTGACCGCGCCAGTCCGGGT	303
Oy	40	ArgctulleasnalaaleaaspyrleuthrAspHisValpheaLaahisLysleu	59
	::::		
Db	304	-----GTCGGGGGGCTACGAGACGGCTGATGGCTGACG-----GTG	342
Oy	60	ProProalIaaspsralIaaspsrglylnala-----	69
		:::::	
Db	343	CCCCGGCGGTGATCGCCGGAACCGGCTGAATCTGATTCAGCAACACTC	402
Oy	70	-----AlaValaspsValhisAsnalaGlnlethrLalauleu	83
		::::: ::::	
Db	403	TTGGGGCAAAACACCCGGGGGATCGGGGTACAGAGCGGATACGGCGAGATGGGGCC	462
Oy	84	ThrlargalaserargleunhisprheuglygluthrProalathrlleaasprphe	103
	:::::		
Db	463	CAGAGCGCGCGGAGATTGTGCGTACGCCGCGGACGCGACGCGCGAGCGGCGGTTG	522
Oy	104	AlaLysalagilyLysleuAsparglyleuAlaThrThrSerglyalaLeu-----	120
	::::: ::::		
Db	523	CTGGCGTTCCAGAGAGCGCGCGAGATG-----ACCAAGCGGGGTGGCTCTCCAGACAG	576
Oy	121	-----ArgalathrProPhealmetAlaser-----LeuGlnllymet	134
	::::: ::::		
Db	577	GCCGCGCGGTCGAGAGAGCGCTCCGACACCGCGCGCGACAGTGTATGAACAATGG	636
Oy	135	GlnProalIaaleasnlysglyAsprPleuProalIaProleuLysProleuthrProleu	154
	:::::		
Db	637	CCCCAGGCGCTGCAACAG-----CTGGCCCAACCCAGACGAGGCAACACAGCT---	684
Oy	155	IleserlyalaleuSerglyalameLaspRlnValGly-----Thr	168
Db	685	---TCCTTCAAGCTGGGTGGCTGTGGAGACGCGTCTCCCGCATCGTCCGCTCAACG	741
Oy	169	LysmetLetaPargalathrthrGlyAspLleuHisLysLysSerAlaserProAsprargleu	188
	::::: ::::		
Db	742	AAACATGGTGCATGGGCCAACACCAACATGTC---ATGACCAACTCGGGTGTGTCATG	798
Oy	189	HisaspalmetalaalaserValLysArgHisSerProserleuLaalarglnValleu	208
	::: ::::: :::::		
Db	799	ACCAACACCTTGACCTGATGTGAAGGGCTTGTCTCCGGCGGCGCCGCCAG-----	852
Oy	209	AsprthrGlyLaIaValGlnThrLysSerAlaArgAsnAlaValaArgthrAlaLeuLa	228
	:::::		
Db	853	-----GCCGTGCAAACCC---GCCGCCCAAAACGGGTCCGG-----	885
Oy	229	ProalaleualaserArgProalalaValGlnGlyAlaValaLaspLeuglyValsermeta	248
	::::: ::::		
Db	886	---GCGATGACCTCGTGGGCGATCGCGTGGTTCCTCGGCTGGGC-----	930
Oy	249	GlyleuLeualaalaasnalaGlyprheGlyAsnargLeuLeuSerValGlnSerArgasp	268
	:::::		
Db	931	GGTGGGGGCGCGCCCACTGGGT-----CGGGCGGCTCGGTCTGGTCTG-----	975
Oy	269	HisGlnargGlyAlaLeuValLeuGlyLeuLysAspLysGlnProLyslaGlnleu	288
	:::::		
Db	976	-----TTG	978
Oy	289	SerglGluAsnAsprTrPleuGlnAlaThrLysAlaIlleLysSerAlaser-----	305
	::::: ::::		
Db	979	TCGGTCCCGCAGGCTGGCGCGCGCCAACCAAGCACTACCCCGCGCGCGCGCTG	1038
Oy	306	-----TyrSerglyalalaLeuAsnalaGlyLysArgmetaLaglyleu	320
	::::		
Db	1039	CCGGTACACACCTGACACAGCGCGGGAAGAAGGGCCCGCGGCAATGCTGGCGGGCTG	1098
Oy	321	ProleuaspmetalaThrAspAlaMetGlyAlaValaArgSerleuValSerAlaser	340

[illegible]


```
;; PRIOR APPLICATION NUMBER: US/09/177,349
;; PRIOR FILING DATE: 1998-10-23
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 2
;; LENGTH: 5036
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
US-09-918-951-2

Alignment Scores:
Pred. No.: 0.00239      Length: 5036
Score: 133.50          Matches: 108
Percent Similarity: 33.41%      Conservative: 43
Best Local Similarity: 23.89%    Mismatches: 182
Query Match: 5.53%             Indels: 120
                                Gaps: 16
                                10

US-09-825-414-7 (1-486) x US-09-918-951-2 (1-5036)

QY 64 SerAlaAspGlyGlnAlaAlaValAspValHisAsnAla----- 76
Db 97 AGCGTCGATGGACAGCGCCGCGTCACGCCGGAGACACTTCTTGATGAAGGAATG 156
QY 77 GlnIleThrAlaLeuIleGlnThrArgAlaSerArgLeuHisPheGlyGlnThrPro 96
Db 157 AAGATGACCTCGCTTATGATTACATCTGAGCTTCCGACGAGACGCCCGCCGG 216
QY 97 AlaThrIleAlaAspThrPheAlaAlaValGlyuLysLeuAspArgLeuAlaThrThr 116
Db 217 TCGTTCGTTGCC-----GCTCCGGGACGGGCGCATGACCACT 252
QY 117 SerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuGlnIleMetGlnPro 136
Db 253 GCCGGGCTGATCGATATGCGCGCCGACCAATCTATCGTGCGGCCCAATGTGTCGG 312
QY 137 AlaIleAsnLysGlyAspThrLeuProAlaProLeuLysProLeuThrProLeuIleSer 156
Db 313 GGTCTGATCTGGGTGCGCGCGACCCC----- 339
QY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlnThrLys-MetMetAspArgAlaThrGln 176
Db 340 -----ATGAGCGGATTCGCGGACGCCGCTCGCCGCTGCGCATGCGTTCGCGAGACGTC 393
QY 176 YAspLeuHisThrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerVal 196
Db 394 GCCA-----ATGTCGGCTCGCGCGGTGACGGCGCGCGCGGTGGCAAGCCTATC 444
QY 196 LlysArgHisSerProSer-----LeuAlaArgGlnValLeuAspThrGlyValAlaVal 214
Db 445 ACGACCGATGTCGGTGGCGCGCTGAGCGACCTGCGTGGGTCTCTGGTGGTCAGGGT 504
QY 214 GlnThrIleThrSerAlaArgAsnAlaValArgThr----- 225
Db 505 GGCCTGCTCTCGCGCGCTCAAGCGGTGTTGCGGCTGACAGTCGCTTGGCTGCCAG 564
QY 226 -----ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAs 242
Db 565 GTCGGTCTGGGTTTACGCGGTGATG-----AGCGCGAG-GTCGGCGCGCTCAGATTGG 617
QY 242 PLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeu 262
Db 618 TGCCTGGATTAGTATTGGACGCGCTGCGTGCAGCGCGGTATAGGC----- 666
QY 262 uSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAsp 282
Db 667 -----TTTGGCGCGCGGGGTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 699
QY 282 sGlnProLysAlaGlnLeuSerGlnGlnLysAsnAspThrPLeuGlnAlaIleLys 302
Db 700 -----CAGCGCGCGCGCGGTATCGG 719
QY 302 sSerAlaSerThrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 322
```

```
Db 720 TGGAGCGCGCGCGGCTATCGCTGCCGGCTGCGGGCTGCGCTTACGCGCAATGCG-- 777
QY 322 uAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuVal----- 336
Db 778 -CAATCGCAATTCGCGCGCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 836
QY 337 -SerAlaSerSerLeuThrGlnAsnGlyLeuAlaAlaGlyLysAlaGlyPheAlaGlyValG 356
Db 837 CCAGCGCGCGCATCGCTAGCCAGATCGGTGTCAGCGCGGTGCGGTGCGGTGCGGTGCGGTGCG 896
QY 356 yLys----- 357
Db 897 CAATGTCAGCGCGCTGACCGGGGTGTCACGACGACGACGACGACGACGACGACGACGACGACG 956
QY 358 -----LeuGlnGlnMetAlaThrLysAsnAl 366
Db 957 CCAGCGCGCGGTGATCCGCACTGAAAGCGCTGCTTGAACGCGCGCTGCTGCTGCTGCTGCTGCT 1016
QY 366 eThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAl 386
Db 1017 GTCGGGCGCGGTGACCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 1076
QY 386 aValPheAlaGlyThrPThrAlaAlaLeuThrThrAspProAlaValLysLysAlaG 406
Db 1077 GTTGGGCTTCGGA---GCGGTGCGGCAACCGCATCTTCAGCGCGCGCGCGCGCGCGCGCGCG 1133
QY 406 uSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyThrValAlaAs 426
Db 1134 TGGCGGTGCGCGCAAGACC-----GAGCGCGCTGCTGAGTGGTGGCGGGGTGCGGCG 1187
QY 426 pGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHisThr 446
Db 1188 GGCAAC-----GCGCGCGGGGTGCGCGGGGCAACGCGCAACATCTGCGGCCAGA 1238
QY 446 rGlyAlaSerLeuArg-----AsnThrValAsn 455
Db 1239 GGGAGCGCGCACTGGGCACTGTCGACAGCGTCAC 1272

RESULT 8
US-09-815-242-4584
; Sequence 4584, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 4584
; LENGTH: 6258
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-815-242-4584

```

Alignment Scores:

Pred. No.:	0.00501	Length:	6258
Score:	131.50	Matches:	125
Percent Similarity:	35.34%	Conservative:	75
Best Local Similarity:	22.08%	Mismatches:	227
Query Match:	5.458	Indels:	139
DB:	10	Gaps:	27

US-09-825-414-7 (1-486) x US-09-815-242-4584 (1-6258)

[illegible]

Db	1294	ATCAACGATAAATCTGCACACATTAGCGACCAAACTTC-----	1332
Qy	265	GlInserTgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro	284
		::: :::	
Db	1333	CTAGATGCTGATGAGCAAAAACGTAATGCTTCAATCAACAGCTGTATTCAAATGGCGAA---	1389
Qy	285	LysAlaGlnLeuSerGluGluAsnAspTTPLeuGluAlaTyrLysAlaIleLysSerAla	304
		::: ::: :::	
Db	1390	---ACAAATTTTAAATAAACAAACTGGACCGCAATACACGCAAAACTCGACGTGCACACAGA	1446
Qy	305	SerTyrSerGlyAlaIleAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet	324
		::: :::	
Db	1447	CTTAATATAT-----GTTAATATAGCGCAAAACATGCATTAAATGTAACCCAAACTTA	1497
Qy	325	AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn	344
		:::	
Db	1498	AATATATGCAAAACACAGCAGATTACAGCAATCAATGAGCGCATCTGATTTAAATCAAAA	1557
Qy	345	Gly-----LeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlu	360
Db	1558	CAAAAAGATGCATTAAAAAGCACAAGCTAATGGTCTCAACGCGTATCTAATGCACACAGAT	1617
Qy	361	MetaIaThrLysAsnIleThrAspProAlaThr-----LysAlaAlaVal	375
		::: ::: :::	
Db	1618	GTA---CAAGCTATGGCACTGAAACGAAACGCGCAATGGCACTTAAACATATGCCATC	1674
Qy	376	SerGlnLeuThrAsn---LeuAlaGlySerAlaIleAlaValPheAlaGlyTTPThrThAla	394
		:::	
Db	1675	GCAGATTAAGCAAGATACGTTACGACACAGTAATATATGTTAACGCGATAGCACTAACAA	1734
Qy	395	---AlaLeuThrThrAspProAlaValLysAlaGluSerPheIleGln-----Asp	411
Db	1735	AATGCTTTACACAACT-----AAAGTTACCATGCTGAAACATATTATTATACGGGTACGCCA	1788
Qy	412	ThrLysSerThrAlaSerSerThrThrGlyTyrVal-----	424
Db	1789	ACGGTGTGTACAAACCTTCAGAAAGTAACGCTGACGTAATCAAGTAACAGCGCGAA	1848
Qy	425	-----AlaAspGlnThrValLysLeuAlaLys-----	433
		::: :::	
Db	1849	CAAGATTAATAATGGTGCAGAAAGATTACGTGTGCAAAAACAAACGCCAATACTGCTATT	1908
Qy	434	-----ThrValLysAspMetGlyGlyGlu	441
		::: :::	
Db	1909	GATGCATTACGCATTAATAATACACCTCAAAAAAGCTTAATTTAAAAACAAAGTGGAGCAA	1968
Qy	442	Ala-----IleThrHisThrGlyAlaSerLeuArgAsnThr	453
		::: :::	
Db	1969	GCCAAATGATTAAGAGACATACAAACGTTTCAAACAAATGAGCAACATGTGACAAATGCA	2028
Qy	454	ValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGlu-----GluGlyGly	471
		::: :::	
Db	2029	ATGAAGAGCTTTAAAGATAGTATTGTCTAACGAAACAAACGTCAAAGCAAGTCAAAATCAT	2088
Qy	472	ThraIaIaSerProSer	477
		:::	
Db	2089	ACAGACGCAACTCCGAAAT	2106
RESULT 9			
US-09-815-242-8512			
; Sequence 8512, Application US/09815242			
; Patent No. US20020061569A1			
GENERAL INFORMATION:			
; APPLICANT: Haselbeck, Robert			
; APPLICANT: Ohlsen, Karl L.			
; APPLICANT: Zyskind, Judith W.			
; APPLICANT: Wall, Daniel			
; APPLICANT: Trawick, John D.			
; APPLICANT: Carr, Grant J.			
; APPLICANT: Yamamoto, Robert T.			
; APPLICANT: Xu, H. Howard			
; TITLE OF INVENTION: Identification of Essential genes in			

FILE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,127
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8512
LENGTH: 17388
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(17388)
US-09-815-242-8512

Alignment Scores:
Pred. No.: 0.0197 Length: 17388
Score: 131.50 Matches: 125
Percent Similarity: 35.34% Conservative: 75
Best Local Similarity: 32.08% Mismatches: 227
Query Match: 5.45% Indels: 139
DB: 10 Gaps: 27

US-09-825-414-7 (1-486) x US-09-815-242-8512 (1-17388)

QY 3 ILeasnAtgATGValGln---GlnProProValThrAlaThrAspSerPheArgThrAla 21
DB 7384 ATTAACATGAGAGTTCAAGTTCGTAATTAACGTAAGTTCGTAACGATTAATAAATGGCAGCA 7443
QY 22 SerAspSerLeuAlaSerSerSerValArgSerVal----- 34
DB 7444 ATGCTACTAATTTAGCTGGTGGTAGCACACAGCAGATTCCTGTGACAGTAATCAAT 7503
QY 35 -----SerSerAspGlnGlnArgGlnLeu 42
DB 7504 GATGGCAGTACTGAGAGATGACAAAGCTCTATTTTCACAAAAGCGGATTAACGTGAGTTA 7563
QY 43 AsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaIleHisLysLeuProProAla 62
DB 7564 ATCAGCAGTAATAATCATTTGATGATCCAGTACAGTACAGTAAAGTAAACCCAGGTACA 7623
QY 63 AspSerAlaAspGlyGlnAlaIleAlaValAspValHisAsnAlaGln-----IleThr 79
DB 7624 ATTACGAGTACAAATTAATGCA-----ATTCATATATGCGCAACAACAATCAATCAAT 7674
QY 80 AlaIleuIleGlnThrArgAlaSerArgLeuHisPheGlnGlyGlnThrProAlaThrIle 99
DB 7675 GCGAAACAGAGCGCA--CAAGTTATTTAATTAATGCGTGCACACCAACAACAAGTT 7731
QY 100 AlaAspThrPheAlaLysAlaGlnLysLeuAspArgLeuAlaThrThrThrSerGlyAla 119
DB 7732 TCTGACGCTTAACATAAGTT-----CGTGTGCACAAACATCAATCAATCAATCA 7779
QY 120 LeuArgAlaThrProPheAlaMetAlaSerLeuLeuGln----- 132
DB 7780 GCTAAAGCA-----TTACTTCAAAATTAAGAAGATATATAGCCAA 7818
QY 133 -----TyrMetGlnProAlaIleAsnLysGlyAspTyrLeuProAla 146

DB 7819 TTAGTAACATCTAATAAATTAACCTACAAAGTCTGTGACCAA----- 7860
QY 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
DB 7861 -----GNACCATCACTACTGTTGATGACGACAAACATATGATTAATCA 7905
QY 167 GlyThrLysMetLeuAspArgAlaThrGlyAspLeuHisThrLeuSerAlaSerProAsp 186
DB 7906 AATCGAAGAAACCGCAAGCAGAAACT-----GAAATTAACGTGACGTCAACGT 7953
QY 187 ArgLeuHis-----AspAlaMetAlaIleSerValLysArgHisSerProSerLeuAla 204
DB 7954 GTTATTTACAAATGCTGTGATGACACTGCACAACAATTTTCAGATGAAGAACTACACGACAA 8013
QY 205 ArgGlnValLeuAspThrGlyValAlaValGlnThrThrSerAlaArgAsnAlaValArg 224
DB 8014 CGCAATGATTTTAACAAATTCAAATTTCCAAAGCTCAACACTTACGT-----GCTGTGAA 8067
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly 244
DB 8068 TCTGTTAAGCAAGTGGC---AATAGTTTAGATGGTGCATGGTACTTACAAACGCGT 8124
QY 245 ValSerMetAlaGlyGlyLeuAlaIleAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
DB 8125 ATCAACAGTAAATCTGGAAACATTAAGCGACGCAAAACTTC----- 8163
QY 265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGlnPro 284
DB 8164 CTGATGCTGTGAGCAAAAACGTAATGCTTACATCAACGTGATCAATATGCCGAA--- 8220
QY 285 LysAlaGlnLeuSerGlnGlnLysAsnAspTyrLeuAlaThrLysAlaIleLysSerAla 304
DB 8221 ---ACAATTTTAATTAACAACACGACGCAACATACGCAAAACTGCAGTCAAGCA 8277
QY 305 SerTyrSerGlyAlaIleAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet 324
DB 8278 CTTAATAT-----GTTAATATGCGCAACATGCAATTAATGCTACGCAAACTTA 8328
QY 325 AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn 344
DB 8329 AATAATGCGAAACAGCAGCATTAACAGCAATCAATGCGCATGCTATTAATCAAAA 8388
QY 345 Gly-----LeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGly 360
DB 8389 CAAAAGATGCATTAAGACACCAAGCTAATGCTGCTCAACGCGTATTAATGCAACAAGT 8448
QY 361 MetAlaThrLysAsnIleThrAspProAlaThr-----LysAlaIleVal 375
DB 8449 GTA---CAACGTAAATGCGACTGAACTGACACGCGCAATGGGCACATTAAACATGCCATC 8505
QY 376 SerGlnLeuThrAsn---LeuAlaGlySerAlaIleValPheAlaGlyTyrThrThrAla 394
DB 8506 GCAGATTAAGACGAATAGCTTACGACAGCAATAATATGTTAAACGCGATACACTAAACAA 8565
QY 395 ---AlaLeuThrThrAspProAlaValLysLysAlaGlnLeuSerPheIleGln---Asp 411
DB 8566 AATGCTTACCAACT---AAAGTTACCAATGCTGTAACATATATTAACGCGTACGCCA 8619
QY 412 ThrValLysSerThrAlaSerSerThrThrGlyThrVal----- 424
DB 8620 ACGGTTGTACACACCTTCGAAAGTACAGCTGCACCTATCAAGTAAACAGCGCGAAA 8679
QY 425 -----AlaAspGlnThrValLysLeuAlaLys----- 433
DB 8680 CAAGATTAATATGTTGACGAAGATVTCGTTGTCACAAACAAACGCCAATACTGCTAT 8739
QY 434 -----ThrValLysAspMetGlyGlyGln 441
DB 8740 GATCATTAACGCAATTAATATACACTCAAAACCTAAATTAAGAAAGACGAGGCGCAA 8799
QY 442 Ala-----IleThrHisThrGlyAlaSerLeuArgAsnThr 453
DB 8800 GCCAATAGATTAGAGACATACAAACTGTTCACAAACAAATGACACAGCATTTGAACATGCA 8859

```

Oy      454 ValmsnsmnuuerrglnatrgpncrGluAlaAspIleGlu-----GluGlyGly 471
      :::: |||||  ||| |||
Db      8860 ATGAAAGCGTTAAAGACGATAGTATTGCTAACGAACACACAGTCAGCAAGTCAAACTAT 8918
Oy      472 ThrAlaAspProSer 477
      ||| |||||
Db      8920 ACAGACCAAGTCGCAAT 8937

RESULT 10
US-09-815-242-8615
? Sequence 8615, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlgen, Karl L.
? APPLICANT: Zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Trawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OR INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815,242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ. ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 8615
? LENGTH: 7035
? TYPE: DNA
? ORGANISM: Staphylococcus aureus
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)...(7035)
US-09-815-242-8615

Alignment Scores:
Pred. No.: 0.00915 Length: 7035
Score: 129.50 Matches: 86
Percent Similarity: 39.55% Conservative: 107
Best Local Similarity: 17.62% Mismatches: 204
Query Match: 5.37% Indels: 91
Db: 10 Gaps: 18

US-09-825-414-7 (1-486) x US-09-815-242-8615 (1-7035)
Oy      8 GlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAla 27
      ||||| ::||| ::| ||| ::::
Db      2290 CAACAATCACAAAGTGTGTCACAAATAAGCTGACTCACAAAGTCTTTACGACGACA 23498
Oy      28 SerSerSerValArgSerValSerSerAspGlnGlnArgLuiLeaSnAlaIleAlaSp 47
      ||| |||::: ::| ||||| ::|
Db      2350 TCTGTGTCATATATGACTTCACATCATGACGTAGTACCTCGAATTCGACAAAGTGTAGC--- 2406
Oy      48 TyrLeuThrAspHisValPheIleAlaIleHisLysLeuProProAlaAspSerAlaSpIly 67
      |||||::||| ||| |||||::| ||| ::::|||
Db      2407 ---CTATCTATTCTGTGAGTGCATCTTAAGTCATTAAAGCAGCATCGAAAGTAAAT--- 2457

```

QY	68	GLIALAIAValAspValHisAsnAlaGlnIleThrAlaLeuIleLeuThrArgAlaSer	87
Db	2458	-----ACGTATCAAGCTTCAACAAGCAAGATTTAGTGAATTCACAAAGCTGA	2505
QY	88	ArgLeuHisPheGluGlyGlu---ThrProAlaThrIleAlaAspRhe-----	103
Db	2506	TCATCAACACGTCAAGGTTCCAGTTAGTAATTCACACATCATTAAGCGATTATTTCGAC	2565
QY	104	---AlaLysAlaGluLysLysLeuAspArgLeuAlaIleThrThrSerGlyAlaLeuArgAla	122
Db	2566	TCACACACTACTGAAATAATAGACAGAGTGTCTCAACAAGATACATCTGATTCATTGGCTACA	2625
QY	123	ThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysSolYasp	142
Db	2626	TCAC-----	2628
QY	143	TripleProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla	162
Db	2629	-----ACATCGTTAAAGCATTCAGTAAGCATAGCATACACAGAAC	2670
QY	163	MetAspGlnValGlyThrLysMetMetAspArgAlaIleThrGlyAspLeuHisTyrLeuSer	182
Db	2671	TTGTCTAG-----TCACAAGCTTTATCAACAAGATACATCCGAT-----TCG	2712
QY	183	AlaSerProAspArg---LeuHisAspAlaMetAlaAlaSerValLysArgHisSerPro	201
Db	2713	GGTATATACATCAACATCGGTAGTGAATGATGATCAATTCATTT-----	2757
QY	202	SerLeuAlaAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsn	221
Db	2758	AGTACATACAGAACTATTAAAGAGATCTCAGGTAGTACATCAGAACTTAACTATTCTAAT	2817
QY	222	AlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal	241
Db	2818	AGCATTTTGCATTCAGTAGAGTGCATCAACAAGAAATTTGGAAATCACAAAGATCATCATA	2877
QY	242	AspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeu	261
Db	2878	TCATTAACTACACATGATTTTCAAATCCATAGTACATCAGAAATCATTAAGCCAT-----	2931
QY	262	LeuSerValGlnSerArgAspHisGlnArgGlyValLeuValLeuGlyLeuLysAsp	281
Db	2932	---TCAACGACACAAAGTGAATTCGTTCTGGATTCATTAAAGTGTACGGGG-----	2979
QY	282	LysGluProLysAlaGlnLeuSerGluLysAsnAspTyrPheGluLysAlaIle	301
Db	2980	-----TCACAAGTGTCTCAACAAGACATCAGCATCATGACATGACACGCTGAATGTATA	3033
QY	302	Lys---SerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320
Db	3034	AGCGATTTCTATGATACAAAGGATGCTATTGTAGCTACATCAGACAGTAAATCAATGTCCCTGA	3093
QY	321	ProLeuAspMetAlaIleThrAspAlaMetGly-----AlaValArgSerLeuVal	336
Db	3094	AGTACTTCGATGACACGCTCTCAGTCAAGTATGATCATCAAAATCATTAAGTCATTCGATA	3153
QY	337	SerAlaSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyLysPheAlaGlyValAlaGly	356
Db	3154	AGTACATCGGATTTCTGACAGTAAGTCAATTATTCTATTAGTACAAAGTCAGCTGTGC-----	3207
QY	357	LysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSer	376
Db	3208	-----TCACAAGACATCAACATCAACAAGATGATTCAGTACGTATGCG	3252
QY	377	GlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeu	396
Db	3253	GAATTCACAAAGTACATCTGGTTCATCAAGCAGCACAAAGTCAATCCGATTCACAAACAGCTATACA	3312
QY	397	ThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr	416
Db	3313	ACG-----TCGTTTATGATTCACACGAGTCAATGACAAA	3345


```

Oy 356 GlyLysLeuGlnIleMetIaThrLysAsnIIeThraSProAlaIeThrLysAlaLa---- 374
Db 1378 GCGAAA---CAAACGGTAGAGTATTCATTAAACAACACATTAAATGAACCAACGTACACGGTTA 1434
Oy 375 -----ValSerGlnLeuThrAsnLeuAlaGlySerFAlaIaIVaIPheIaGLYTrP 391
Db 1435 GACATGAATTAATTCACAGCACACACAAATGTGGAAGTGTTAATACAGTTAAACC----- 1488
Oy 392 ThrThraIaaLeuThrThraSProAlaVaLYsLYsaLIaGluseRPhelIeGInaSP 411
Db 1489 -----AAAGCGCAACAATAGATGGCTGCTATGGGTCAATTGAAACATCATATTGGTAT 1542
Oy 412 ThrValLyseRThraIaseRserThrrHglyTYr----- 423
Db 1543 -----AAAGACAGCAGCTTACCAAAAGTCCAAATTCACAGATGCTGATGATGCTAACGA 1596
Oy 424 ---ValAlaSPGIthrThraLYsLeuAlaLYsthrVaI-----LysaSPmeIGLYgLY 440
Db 1597 ACTGCTATTCTCAGACAGATAAATGCACAGCACTATTGTTAAATPAACAGAGTGGCGGT 1656
Oy 441 -----GluAlaIleThriSthrgLYaIaser 449
Db 1657 AATACACTTAAGACAGATGTGTAAGAAGCAATGCCAAGCTGTACACAAACAAATCTGCA 1716
Oy 450 LeuArgAsnThraValasn----- 455
Db 1717 TTAAACGCTATTCAAACTTAGTCGTGCGCAAAACGCTGCTAACACAGCATTTACAAAT 1776
Oy 456 -----AsnLeuArgGInaIgrProIlaIatrgGuaIaSPllIGlUGLYgLY 471
Db 1777 GCTTCGAGCTTAATACAAAAACAAAAAGAACATTAAAGACAAAGTAAACAAAGTGCAGGA 1836
Oy 472 ThraIaIaseRProserGluIue 479
Db 1837 CGGTATCTCGACCAAAATGTGTT 1860

RESULT 12
US-09-815-242--8898
US-Sequence 8898, Application US/09815242
Patent No. US20020061569X1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes In
FILE REFERENCE: Prokaryotes
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8898
LENGTH: 18846
TYPE: DNA

```

Db	3409	GGTGCACACAGCTGCTGCTGCTAAATCATCAGAAACTGCAGAAAGCCACACAGATTAATTAATAC	111	3468
QY	260	ArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeu	111111	279
Db	3469	GCAATGCATAGCTTTCACAAAT	111111	3495
QY	280	LysAspLysGlnProLysAlaGlnLeuSerGlnGluAsnAspThrPheGluAlaTyrLys	111111	299
Db	3496	NATGAT	111111	3496
QY	300	AlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLys	111111	315
Db	3547	AGTAGAATACGCT	111111	3600
QY	316	ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeu	111111	335
Db	3601	AAAGCTAGCGTCACAAATGTAGACAAAGCA	111111	3654
QY	336	ValSerAlaSerSerLeuThrGlnGlnGlnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal	111111	355
Db	3655	GGAACACAGTACACAGACGCGCTTGAAAGCGTATGCGCAAAATTAATGAGCTAAAGCAGCT	111111	3714
QY	356	GlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAla	111111	374
Db	3715	GGGAAA	111111	3771
QY	375	ValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyr	111111	391
Db	3772	GACAAATGAAATATACACAAACAAATGTGGAAGGTGAATATACAGTTAAAGCC	111111	3825
QY	392	ThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAsp	111111	411
Db	3826	AAAGCCACAAATTAGCTGCTATGCGTCAATTAGAAACATCATTTCCGTGAT	111111	3879
QY	412	ThrValLysSerThrAlaSerSerThrThrGlyTyr	111111	423
Db	3880	AAAGACACGACGCTTACAAAGTCAAAATTAACAAGATCGTATGCTGTAACGA	111111	3933
QY	424	ValAlaAspGlnThrValLysLeuAlaLysThrVal	111111	440
Db	3934	ACTGCTATTATTCAGACAGTAAATGACGACGACAACTATTTAATTAACAGCTGGCGGT	111111	3993
QY	441	ValAlaSerSerThrAlaSerSerThrThrGlyAlaSer	111111	449
Db	3994	AATACACCTAAAGCAGATGTGGAAGACAGATCAAGCTGTATACACAAAGCAAAATACTGCA	111111	4053
QY	450	LeuArgAsnThrValAsn	111111	455
Db	4054	TTTAAACGATTCACAAACTTAGATCGTGGGAACAGCGTCTACACACAGCATTTACAAAT	111111	4113
QY	456	AsnLeuArgGlnArgProAlaArgGluAlaAspIleGlnGlyGly	111111	471
Db	4114	GCTTCGACACTTAATACAAACAAAGAGCACTTAAAGCACACAGTAACAACTGCAGCA	111111	4173
QY	472	ThrAlaAlaSerProSerLeuIle	111111	479
Db	4174	CGTGTATCTGCAGCAAAATGTGT	111111	4197
RESULT 13				
US-09-287-849-25				
Sequence 25, Application US/09287849				
Patent No. US2002009459A1				
GENERAL INFORMATION:				
APPLICANT: Reed, Steven G.				
APPLICANT: Skelky, Yasir A.W.				
APPLICANT: Dillon, David C.				
APPLICANT: Alderson, Mark				
APPLICANT: Campos-Neto, Antonio				
APPLICANT: Corixa Corporation				
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens				
TITLE OF INVENTION: and their uses				
FILE REFERENCE: 014056-00902005				

[illegible]

Db 742 AACATGTCATGATGCCAACACACATGTCG---ATGACCAACTCGGTCGTGCATG 798
Qy 189 H1aSPALmeTAlaIaSerValLysArgH1sSerProSerleuAlaArgInValLeu 208
Db 799 ACCAACACCTTGAGCTGATGTTGAAGGCTTTGCTCGGCGGCGCCAG-----852
Qy 209 AspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla 228
Db 853 -----GCCGTGCCAACCC---GCCGCCCAAAACGGGGTCCGG-----885
Qy 229 ProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValSerMetAla 248
Db 886 ---CGCATAGCTCGCTGGCGCATCGTCGTCTTCGGGTGGCG-----930
Qy 249 GlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAsp 268
Db 931 GGTGGGTGGCGCCCAACTTGGGT-----CGGGCGGCTCGGTGGGTGG-----975
Qy 269 H1aGlnArgGlyAlaLeuValLeuGlyLeuLysAspLysGlnProLysAlaGlnLeu 288
Db 976 -----TTG 978
Qy 289 SerGlnGluAsnAspTrpLeuGlnAlaTyrLysAlaIleLysSerAlaSer-----305
Db 979 TCGGTCCCGAGGCTGGCGCGCCCAACAGCAGTCAACCCCGCGCGCGCGCGCTG 1038
Qy 306 -----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 1039 CCGCTGACCACTGACACAGCGCGCGGAAAGAGGCGCGCGAGATGCTGGCGGCGCTG 1098
Qy 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 1099 CCGGTGGCG-----1107
Qy 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLeu-----358
Db 1108 ---CAGATGGCGCCCAAGGCGCGGTGGTGCATCAGTGTGCTGGCTGCGCGCG 1161
Qy 359 -----GlnGluMetAlaThrLysAsnIleThrAspProAla-----370
Db 1162 CGACCTATGTATGCCCGCATTTCTCCGGCAGCGCGGATATCCCGCGCGGCTGTGTCG 1221
Qy 371 -----ThrLysAlaAlaValSerGln 377
Db 1222 CAGGACCGGTGCGCGACTTCCCGGCGTCCCGCTGACCCGTCGCGGATGTCGCCCA 1281
Qy 378 Leu-----ThrAsnLeuAlaGlySerAlaAlaValPheAla 389
Db 1282 GTGGGCGCACAGGTGTCAACATCAACACCAACTGGGCTACAAACAGCGGTGGCGCC 1341
Qy 390 GlyTrpThrThrAlaAlaLeuThrThrAspPro-----400
Db 1342 GGG-----ACCGGCATCGTATCGATCCCAACGGTGTCTGCTGACCAACACAC 1392
Qy 401 -----AlaValLysLysAlaGlnSerPheIleGlnAspThrValLysSerThrAla 417
Db 1393 GTGATGGCGGCGCCACACGATCAATGCGTTCAGCGTGGCGTCCGCCCAACACTACGGC 1452
Qy 418 SerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLys-----433
Db 1453 GTCGATGTGGTGGTATGACCCGACCGAGATGCGCGTGTGACGTCGCGCGTGGCC 1512
Qy 434 -----ThrValLysAspMetGlyGlyAlaIleThrHisThrGlyAlaSer 449
Db 1513 GGTGGCTCGCGTGGCGGCGGATCGGTGGCGGCGGTGGTGTGAGCCCGCTGCGCG 1572
Qy 450 LeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGlnAlaAspIleGlnGlu 469
Db 1573 ATGGCAACAGC-----1584
Qy 470 GlyGlyThrAlaAlaSerProSerGlnIleProPheArg 482
Db 1585 GTGGGCAAGGCGCGCCCTGCGGTGTCGCGAGG 1623

RESULT 14
US-09-860-846-32/C
Sequence 32, Application US/09860846
Patent No. US20020164742A1
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
TITLE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600,438US1
CURRENT APPLICATION NUMBER: US/09/860,846
CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32
LENGTH: 11220
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-860-846-32
Alignment Scores:
Pred. No.: 0.0267 Length: 11220
Score: 127.50 Matches: 113
Percent Similarity: 36.10% Conservative: 61
Best Local Similarity: 23.44% Mismatches: 190
Query Match: 5.29% Indels: 122
Gaps: 18
US-09-825-414-7 (1-486) x US-09-860-846-32 (1-11220)
Qy 14 AlaThrAspSerPheArgThrAlaSerAspAlaSerLeu-----AlaSerSerSer 30
Db 8351 GCGGGCGAGCGCACAGAGAGCGTCCGCGGGCGGTGCGCCCGCTGACGAGCGCTCC 8292
Qy 31 ValArgSerValSerSerAspGlnGlnArgGlnIleAsnAlaIleAlaAspTyrLeuThr 50
Db 8291 ACCAGGTCCCTTAAGTCCGGGAAAGAGACAGACTCGGTGCCACCGCGCCAGCGCGCG 8232
Qy 51 AspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyAlaAla 70
Db 8231 GCGAGACCGTCGCGCGTCC-----CGGCCAGCAGCACCGCGCGGCGAGGCGCTGC 8181
Qy 71 Val-----AspValHisAsnAlaGlnIleThrAlaLeuIleGlnThrArgAla 86
Db 8180 GCGGACCGTCCCGAGCGCGGTCCACTCCAGCAGGTGCAGCGCTCCAGAGTCCGGTCCGTG 8121
Qy 87 SerArgLeuHisPheGlnGlyIleThrProAlaThrIleAlaAspThrPheAlaLysAla 106
Db 8120 AAGCGCCACGTCGCGGGGTGACGCGGCGACGAGTCCGTAGGAGTCCGCGGACACCC 8061
Qy 107 -----GluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThr 123
Db 8060 GCGTGGCGGAGAGAGTCCGGGCGGTCCAGGACACGCTGTCGGCGCGGCGGCGGCGAG- 8002
Qy 124 ProPheAlaMetAlaSerLeuLeuGlnTyrMetGln-----ProAlaIleAsnLysGly 141
Db 8001 CCGCAGCGCGAGGCGCGGTGCGCGCCGACCGCGTACAGAGATCCCGCTCCAGCGAACCGCG 7942
Qy 142 AspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGly 161
Db 7941 AGCGGTGTCGCCCGCGGAAGCGCCGCGCGC-----ACCGCGCGCTGCACGGC 7891
Qy 162 AlaMetAspGlnValGlyThrLysMetMetAspArgAla-----ThrGlyAspLeu--- 178
Db 7890 GCGGTGAGAGAGCGCGGCGGAGAGCGGAGCGGAGCGCGCGCTTGACACCGCGACCTCGGC 7831
Qy 179 -----HisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMet 192
Db 7830 CCGCAGGCGCACGTCGCGCGAACACCTGTCGCCACGCGCCACAGACACACGAGCGCTG 7771

OY	193	-----AlaAlaSerMetLysArgHisSerProSerLeuAlaArgGlnValLeu	asp	209
		:		
Db	7770	GAAGAGGGGGCCGTAACCGTA	-----GCCGTTCCGCCGAGACGCTTGTA	7726
OY	210	ThrGlyAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla	Pro	229
Db	7725	CAGACCGTCCACGTCACCGGCTCGGGCGCCGCGG	-----	7690
OY	220	AlaLeuAlaSerArgProAlaValGln	-----GlyAlaValAspLeuGly	244
Db	7689	-----CGGCCAGGCTCCCGGGTCCGACGGGGGGCGTCCGTCACAGGGGC	7642	
OY	245	ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuSerVal	264	
Db	7641	GGCCAGCACACCGGTGGCTGGCCGCTCCACTCCGCTCCCGCGGCTCTCCGGGTG	7582	
OY	265	GlnSerArgAspHisGlnArgGlyAlaValLeuVal	-----	276
		:::		
Db	7581	CGCGTAGAGCCCGAAGGTATGACGACGCCCGGACTGTGCTCGCGCGAGGACAGCTGCAC	7522	
OY	277	-----LeuGlyLeuAspPylsGlnProPylAlaGlnLeuSerGln	291	
Db	7521	ACGGACCGCGCCACGACGAGGGGACACAGACCGCGCGTGAAGGTGAGTCTCTCACAG	7462	
OY	292	-----AsnAsp-TrpLeuGlnValTyrAlaIleLeuSerAlaSerTyrSerGlyAl	309	
Db	7461	ATCCGACCGACCTGCTGCCCGCTGGGACGACCTCCACGACGCCGCTCCCGGACG	7402	
OY	309	AlaLeuAsnAlaGly	-----	314
		:::		
Db	7401	CAGCAC-GGTGCCGGCCACCGCTGGTCCGCGACGAGGGTGGTACGAGGAGAGAGC	7343	
OY	315	-----LysArgMet	317	
Db	7342	TCCCGGTGACGACGACCGCTCGGAGTCCGCGAGCGCCAGCGCCCGCCGACACCGGGT	7283	
OY	317	TAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuVal	337	
		:::		
Db	7282	GCTCGCGCGCCCGACAGACCGCGGAGGTATGTACACGGCGGCGG-AGAAGTCCGGGCTGC	7224	
OY	337	ValSerSerLeuThrGlnAsnGlyLeuAlaLeuAla	-----GlyGlyPheAlaGly	355
Db	7223	GGCCAGTAGCGCTCGCTGTGAAGCGGTAGTGGCAGATCGGGGGGTGCTGTG-GT	7165	
OY	355	IGlyLysLeuGlnGlnMetAlaThr	-----	363
Db	7164	GGGAGGAGAGAGGCCACGTCGACGGTGAAGCCGTTGGCCAGGCCCTCGCGAGGAGGT	7105	
OY	364	-----LysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl	382	
		:::		
Db	7104	GGTAGGCGGCTGCTCCGCGCTTGTCACGCGCGAGGGTCCGAGTCCGGTACCGTCTC	7045	
OY	382	aglySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAla	402	
		:::		
Db	7044	GGCGACGCGCATGGTGAAGCGGGGTGG-----GCCGTGACCTCGACGAAGTGGGT	6994	
OY	402	ILys 403		
Db	6993	GAGG 6990		
RESULT 15				
US-09-861-289-32/c				
Sequence 32, Application us/09861289				
Patent No. US2002010897A1				
GENERAL INFORMATION:				
APPLICANT: Sherman, D.H.				
APPLICANT: Liu, H.				
APPLICANT: Xue, Y.				
APPLICANT: Zhao, L.				
TITLE OR INVENTION: DNA encoding methymycin and pikromycin				
FILE REFERENCE: 600.4380S1				
CURRENT APPLICATION NUMBER: US/09/861,289				

```

US-09-825-414-7 (1-486) x US-09-861-289-32 (1-11220)

CURRENT FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: 09/105,537
PRIOR FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 32
LENGTH: 11220
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Alignment Scores:
Pred. No.:          0.0267          length:          11220
Score:              127.50          Matches:         113
Percent Similarity: 36.108         Conservative:    61
Best Local Similarity: 23.444       Mismatches:     190
Query Match:        5.298          Indels:         122
DB:                  10            Gaps:           18

```

US-09-825-414-7 (1-486) x US-09-861-289-32 (1-11220)

QY 14 AlAThrAspSerPheArgThrAlaSerAspAlaSerLeu-----AlaSerSerSer 30

Db 8351 GCGGGGCAAGCCACCAGGACGGTCCGCGGGGGCCGGGGTCTCGCCGCCCGGTCGACGGCCTCC 8292

31 ValArgSerValSerSerArgInGlnArgGluLeuAsnAlaLeuAlaAspTyrLeuThr 50

DB 8291 ACCAGGTCCTAAGGTCGGGAAGGACAGGACCTCGGTGCCACCGGCGCAGCGCGG 8232

```

cy      |  |||:::  |||:::  |||:::  |||:::  /
jt asphnswalfnehaiahaahsbysbeufiofiowiaaspsbeiaiaaspolymiaiaia  /

```

[illegible][illegible][illegible]

1. $\frac{1}{2}$

[illegible][illegible][illegible][illegible]

145 ASRTnIeuprC]a prol,eul,xsProl,eulThrProl,eul]eSerg]va]aI,euSerg]v 161

Dp 7941 AGCCGCTGCCCGCCGAACGCCCGGGCCGC-----ACCGGCCCTGCACGGC 7891

QY 162 ALamETaSPGI nValGI VThrLysMetMetaspARqAla-----ThrGI yASpLeu--- 178

Db 7890 GGCCTGAGCAGCGCCGGTGAAGCCGACCGCGCCCTCGGCACCGCGACCTGGC 7831

QY	179	-----Histyrl	192
		euSerProaspa	
		glLeuHisasp	
		lasermet	

Db 7830 CGG CAGGGCCACGTCGGCGAACACCTCGTCGCCACGCCGCCAGACACCAACGAGCGCCCTG 7771

```

QY 193 -----AlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuasp 209

```

Db 7770 GAAGAGGGGCGCTAGCCGTA-----GCCGTTGCCGCGGAAGCGCTCGTA 7726

210 ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229

Db 7725 CAGACCGTCCACGTCACCGGCTCGGGCGCCCGGGCG----- 7690

QY	230	AlaLeuAlaSerArgProAlaValGln-----	GLYAlaValAspLeuGly	244

DB / 089 -----CGGCCAGGCTTCGGGTCCGGACGGGGCGGTGCGGATCCGGACGGGC / 042

```
OY 245 ValSerMetAlaGlyLeuAlaIAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
Db 7641 GCCAGCACACCGGTGGCGCGCTCCACTCCGCTCCCGCGCGCTCCGGGTG 7582
OY 265 GlnSerArgAspHisGlnArgGlyAlaLeuVal 276
Db 7581 CGCGTAGAGCCCGAGTAGACAGCCCGGACTCGTCGCTCCGCGCGAGACAGCTGCAC 7522
OY 277 -----LeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGlu 291
Db 7521 ACGACCCGCCACACGACGCGGAGCGCGCTCGAGGGTAGACTCTCCGACACAG 7462
OY 292 -----AsnAsp-TrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAl 309
Db 7461 ATCGAACCGACCTGTCTCCCGCTCGAGACGACAGCTCCACGACGCGTCCGCGCAG 7402
OY 309 AlaLeuAsnAlaGly 314
Db 7401 CAGCAC-GGTGCGCGCACCGCGTGTCCGCCACCCAGGGTGCCTACGAGGAGAGGC 7343
OY 315 -----LysArgMet 317
Db 7342 TCCCCGTAGAGCAGCAGCGCTCGAGTCGCGAGCGCACCGCCGCGACACCGGAT 7283
OY 317 tAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSe 337
Db 7282 GCTCGCGCGCGCGAGACCGCGGAGGTGATCTCACCGCGCGG-AGAAGTCGGGCTGC 7224
OY 337 rAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAla-----GlyGlyPheAlaGlyVa 355
Db 7223 GGCCAGTAGCGCTCGCTCGAAGGCGTAGTGCGCAGATCGGGTGGGTGCTGTG-GT 7165
OY 355 lGlyLysLeuGlnMetAlaThr 363
Db 7164 GGGGAGGAGAGAGCGCCAGTCGAGCGTAGCGCGTGGCCAGGCTCGCGCAGGAGAGT 7105
OY 364 ---LysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl 382
Db 7104 GGTGAGCGCGGTCTCCGCGCTGTTCACGGGAGGGTCCGAGATCCGGTACCGTCTC 7045
OY 382 aGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaVa 402
Db 7044 GGGCAGCGCCATGTGTGAGGACGGGGTGG-----GCGTGCACCTCGACGAGTGGGT 6994
OY 402 Lys 403
Db 6993 GAAg 6990
```

Search completed: January 31, 2003, 07:21:46
Job time : 138.921 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:25:34 ; Search time 3127.71 Seconds
(without alignments)
4531.450 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407
Sequence: 1 MHINSAQPPGVAMSPRT.....EEGGSARSRETPQLRRL 487

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09825414/runat_28012003.161741.26085/app_query.fasta.1.1294
-DB=GenEmbl -QEMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=txt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414.ecgn.1.1.4707_gunat.28012003.161741.26085 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DLOGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: GenEmbl:*
2: gb_ba:*
3: gb_hg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pa:*
8: gb_ph:*
9: gb_pl:*
10: gb_pr:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vl:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_or:*
23: em_ov:*
24: em_pat:*
25: em_ph:*
26: em_pi:*
27: em_ro:*
28: em_sts:*
29: em_vl:*
30: em_hg_hum:*
31: em_hg_inv:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pln:*
35: em_hg_rod:*
36: em_hg_mam:*
37: em_hg_vrt:*
38: em_sy:*
39: em_hggo_hum:*
40: em_hggo_mus:*
41: em_hggo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1869.5	77.7	52498	1	AF232004	AF232004 Pseudomon
2	1591.5	66.1	1834	1	AF458051	AF458051 Pseudomon
3	460	19.1	208050	1	AL646083	AL646083 Ralstonia
4	160.5	6.7	1358	3	AY094894	AY094894 Drosophila
5	160.5	6.7	163709	3	AC108877	AC108877 Drosophila
6	150.5	6.3	203050	1	AL646078	AL646078 Ralstonia
7	149.5	6.2	10029	1	AE008061	AE008061 Agrobacte
8	149.5	6.2	12020	1	AE009096	AE009096 Agrobacte
9	149	6.2	12313	1	AE008285	AE008285 Agrobacte
10	149	6.2	12568	1	AE009328	AE009328 Agrobacte
11	146.5	6.1	301769	3	AE003482	AE003482 Drosophila
12	146	6.1	306250	1	SME591788	AL591788 Sinorhizo
13	145	6.0	13040	1	AE005769	AE005769 Caulobact
14	143.5	6.0	6413	1	AF677776	AF677776 Abiotroph
15	143	5.9	132457	9	AC016584	AC016584 Xanthom
16	142	5.9	16675	1	AE011814	AE011814 Xanthom
17	142	5.9	138685	2	AC017582	AC017582 Drosophila
18	142	5.9	166249	3	AC093502	AC093502 Drosophila
19	142	5.9	185404	3	AC104703	AC104703 Drosophila
20	142	5.9	341319	3	AE003536	AE003536 Agrobacte
21	141	5.9	9083	1	AE008286	AE008286 Agrobacte
22	141	5.9	11449	1	AE005078	AE005078 Halobacte
23	140.5	5.8	291150	1	AP003135	AP003135 Staphyloc
24	140.5	5.8	342600	1	AP003363	AP003363 Staphyloc
25	140	5.8	1512	1	PF8297534	AJ297534 Pseudomon
26	140	5.8	7100	6	AX004713	AX004713 Sequence
27	140	5.8	134416	7	AF020713	AF020713 Bacteriop
28	140	5.8	213680	1	BSUB0012	Z99115 Bacillus su
29	139.5	5.8	1473	1	AB073918	AB073918 Campyloba
30	139.5	5.8	8923	1	AB015223	AB015223 Staphyloc
31	139.5	5.8	304050	1	AP004829	AP004829 Staphyloc
32	138.5	5.8	1473	1	AB080202	AB080202 Campyloba
33	138.5	5.8	3433	1	SDY243529	AJ243529 Streptococ
34	138.5	5.8	12312	1	AE012307	AE012307 Xanthomon
35	138	5.7	3831	6	AX066471	AX066471 Sequence
36	138	5.7	6136	6	AX066467	AX066467 Sequence
37	138	5.7	8979	6	AX122823	AX122823 Sequence
38	138	5.7	10635	1	AE011815	AE011815 Xanthomon
39	138	5.7	13431	1	AE000234	AE000234 Escherich
40	138	5.7	16902	1	D90775	D90775 E.coli geno
41	138	5.7	18700	1	D90774	D90774 E.coli geno
42	138	5.7	47233	2	AC017267	AC017267 Drosophila
43	138	5.7	156251	2	AC107325	AC107325 Drosophila
44	138	5.7	194657	2	AC009376	AC009376 Drosophila
45	138	5.7	287649	3	AE003516	AE003516 Drosophila

RESULT 1

ALIGNMENTS

AF232004/c 52498 bp DNA linear BCT 05-MAR-2001
LOCUS
DEFINITION Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity
island, complete sequence.
ACCESSION AF232004 AF061028 AF061029 AF232006 LA1861
VERSION
KEYWORDS AF232004.3 GI:13325077
SOURCE
ORGANISM Pseudomonas syringae pv. tomato.
Pseudomonas syringae pv. tomato
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas
1 (bases 25494 to 29778)
Preston,G., Huang,H.C., He,S.Y. and Collmer,A.
The HrpZ proteins of Pseudomonas syringae pvs. syringae, glycinea,
and tomato are encoded by an operon containing versinia ysc
homologs and elicit the hypersensitive response in tomato but not
soybean
Mol. Plant Microbe Interact. 8 (5), 717-732 (1995)
JOURNAL
MEDLINE
PUBMED 96025089
REFERENCE 7579616
AUTHORS
TITLE 2 (bases 22134 to 25847; 29687 to 32670)
Deng,W.L., Preston,G., Collmer,A., Chang,C.J. and Huang,H.C.
Characterization of the hrpC and hrpS operons of Pseudomonas
syringae pathovars syringae, tomato, and glycinea and analysis of
the ability of hrpF, hrpG, hrpC, hrpT, and hrpV mutants to elicit
the hypersensitive response and disease in plants
J. Bacteriol. 180 (17), 4523-4531 (1998)
JOURNAL
MEDLINE
PUBMED 98389667
REFERENCE 9721291
AUTHORS
TITLE 3 (bases 31672 to 51723)
Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and
Collmer,A.
The Pseudomonas syringae pv. tomato HrpM protein has domains
similar to harpins and pectate lyases and can elicit the plant
hypersensitive response and bind to pectate
J. Bacteriol. 180 (19), 5211-5217 (1998)
JOURNAL
MEDLINE
PUBMED 9748456
REFERENCE 98422476
AUTHORS
TITLE 4 (bases 901 to 22404; 31672 to 51723)
Alfano,J.R., Charkowski,A.O., Deng,W.L., Badel,J.L.,
Penicki-Ocvietje,T., van Dijk,K. and Collmer,A.
The Pseudomonas syringae Hrp pathogenicity island has a tripartite
mosaic structure composed of a cluster of type III secretion genes
bounded by exchangeable effector and conserved effector loci that
contribute to parasitic fitness and pathogenicity in plants
Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)
JOURNAL
MEDLINE
PUBMED 20243785
REFERENCE 10781092
AUTHORS
TITLE 5 (bases 1 to 52498)
Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Pseudomonas syringae pv. tomato DC3000 hrpL through hrpC
unpublished
6 (bases 1 to 52498)
Alfano,J.R. and Collmer,A.
Direct Submission
Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland
Parkway, Las Vegas, NV 89154, USA
7 (bases 1 to 52498)
Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Direct Submission
Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334
Plant Sciences Bldg., Ithaca, NY 14850, USA
8 (bases 1 to 52498)
Sequence update by submitter
Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Direct Submission
Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334
Plant Sciences Bldg., Ithaca, NY 14850, USA
REMARK
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

/organism="Pseudomonas syringae pv. tomato"
/strain="DC3000"
/db_xref="taxon:323"
/note="exchangeable effector locus flanking the hrp/hrc
cluster
pathovar: tomato"
complement(908..1972)
/gene="queA"
complement(908..1972)
/gene="queA"
/codon_start=1
/transl_table=1
/product="QueA"
/protein_id="AA071483.1"
/db_xref="GI:8037772"
/translation="MRVADFTFELPSLIRHPLAERSSRLTLDPGALAHROPT
DLLEHRSGLMVFNNTRVYPAFLFGKAGSGLEILEIVELDLSHYLAHRAKSPK
PGSSITLIDSGGEAEVVARHDLFLREAEVLPDLIDVGHMPLPYITDRPDEADRER
YQVYVAGRAGAVAPTAGLHFDQPLMEAIKAVETAFTLVHAGAGTPOVREQLIED
HHHSEWLEVSQDVDAVVAACRARGRVIAVGTTSVLSLSAARDGLKPFSGDTDF
ITYGRPHVVDALVTNHLPESTLLMLVSAFAGVPETMAYAAIEHGYRFPYGDAM
FTRNPAPTAPQESAPEDHA"
2030..2177
/product="tRNA-Leu"
2343..2753
/note="ORF4"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AA071484.1"
/db_xref="GI:8037773"
/translation="MNKIVYKAYFKPIGEVSVKVPTEIRKFGDEIKMKETQM
OQGWMSQCIDGEBLSKDVDAVAQLMADGEIQTVLPILISGAYDAALKRYEIRHNR
TELSPGOSYVFGYGSFTGCVTLVAKKFGSSMS"
complement(2908..3324)
/note="ORF3"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AA071485.1"
/db_xref="GI:8037774"
/translation="MRAVKNLTKAGISFLALITIGSLPAPVYVNDCLDNDNSTGAT
CGGNDKLDNDVNTIDAFAFGNDMDMDHTDFAFGGNDKLDNDHTDFAFGGNDK
LDNDKTDAAFGVNRDLDDNDNTNNGTSPSAKK"
complement(3939..4664)
/note="ORF2"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AA071486.1"
/db_xref="GI:8037775"
/translation="MOKTLMALAFAMLAGCGVSGPAPGSDIOGAQAEKTPVKLND
AYTSKRDVLAETARNTSVNKGOLIDIVSAGFLGIPYRBNMLVGSANVPEQVDR
GLDGFATLDVYEAERFSTSOODFYRNLYVQRYKGGVDYFINKHFTPDNAYGTATPPA
DDITTAQISPGAVSVKRLNERAKGKTYLPPLPYVERSMYTIPLSLVDSQVSHLRGTD
YIGTYPSASAGCDTRRFLYRDG"
5571..6410
/gene="tnpA"
5571..6410
/gene="tnpA"
/note="truncated at the amino terminus"
/codon_start=1
/transl_table=1
/product="tnpA"
/protein_id="AA071487.1"
/db_xref="GI:8037776"
/translation="WGFRLPAPYPLAMLRITHLMNPNFVSDPAMEALVETTLRQFA
GLSIDRLADETTLINFRRLKEHLAAGLIVTINGLVGDRGLMROGMVADATLTHAP
SSTRKNKGKDEPMHQTKGNQYFFGKRAITGVADBSGLVHSVGTAAVADYQVDO
ILHSEERYVSGDGYGVDRRAEHODOMSTYIAPRSRYKSGEKSILARVRYKTEF
TKAOLRAKVHPLRVYIKRQGYIKVRRGLAKNTAQAATLFAISNLMMVRRKLANGE
VRL"
complement(6656..8056)

```
/note="orf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF71488.1"
/db_xref="GI:803777"
/translation="MRPVGGAPGYPTTEAEERPTAQAAGNDARRSSQASSPAVY
APETPMGLDTRKRPAGRPDMKVENIRKITEGPERGKDGKHKRRRKRDLAAGSSHYH
GGOSVASTSASAKSLADDTNFKASDLAELRWCSPPRYALASKAAGKSQISANY
VSLQEGKHALEORLEAGLADVVYSEGRDHILNLTLEMDSCLGSGILWAPD
SNDKRLKLAKARYPDDENAOKLPELAPLTKKSKDSLSGMEELLADAPLIGEHNS
TSSKRELINNNKSLKASGVTLFMELHCAESHDALNNYLSAPKSGPPARALYDLA
OSQHOAPEELTKYNETTLVEAKHACLRYVSLDPTSYVAPEKATIKRAQAMNYLA
AEKIRLSPEKKWAFVGATHTATSGDGPGLAELHGVASLVIDDLGLKSRATVIDNK
NYGKLNPDVRLSYK"
complement(8131..10473)
/gene="hrpk"
/complement(8131..10473)
/gene="hrpk"
/codon_start=1
/transl_table=11
/product="hrpk"
/protein_id="AAF71489.1"
/db_xref="GI:803778"
/translation="MRISSPFVINYOPTPEGLALAVESPLAKALPTPVGGGAGVQ
FGOPAGNTOGAPTGAEOPTASSILALLDGGGAANANVNOVOPOPCADSPAEAPT
EATAPAPVEOVAAPTNTDATTASNAKPAADAPLNDSEYSPALKKMEPMVHLRP
EERDQAKELNRPITAAAMMARDDCGPKAKMDFINANPLATVADVAGGGNADKIT
NKDLTKFKNMKADSDADKDLANTKONPADPQSLMVSAAVMRANPLATLADP
HHAVDAADKTVEGDNVADDELGLAOSNPGSGLAKOSCTWSQPGFQGVDEAGMSG
RKKAHSPDKMEADKNSLEWIKKSAPTMGQFASMLSDSATLNAAGIDISKLQVDF
DKPKSYSGAOKMAVWVKLOOTQOQSYIAGRSIRNDTKTEGNDRISSOLDADVOAYL
NKSJPEOERNLVBSDAOLOKAVNEOTKRVNSGOALOTMDKADRAVHNHNANYSGA
ISGSAOQLOKDLFPDSKVPPTNVYLENKDLDKITSTVTNFSSEGALKQLGOK
KSDASQSLQTDNOKAIDSVLPDPTENQHNWNSLTGLDYSKRRKLEKSDG
GGPMLTQOLAAGVGKALNSIVGFASVSLIAHQDLGATQALYDSTRMGAELKKG
IDAGAKMMGREGASAGLGRIGCOMIGRAVGNVAGEATGILAAAGALIPVGMADGA
LAVFGISMITDAVKKHKAOKAFDHNVDPVLDQFCIAKAN"
complement(110560..11114)
/gene="hrpl"
/complement(10560..11114)
/gene="hrpl"
/note="alternative sigma factor"
/codon_start=1
/transl_table=11
/product="hrpl"
/protein_id="AAG33875.2"
/db_xref="GI:13325078"
/translation="MFQIVILDSSTQPPQSSSAGIRQMTADQIOMLRAFIQKRVMP
DDVDIQCVELEALRNEHKFOHASKPQTMCGTALNIRNHFRRMYRQPOQESWED
VHSELEGHGDSHOVEGHROLARVQALDCLPSNNQKYLEVSLMDGNYQETANSLSY
PIGTVRSRLSARVOLKQIDPFA"
11358..12464
/gene="hrpj"
11358..12464
CDS
gene
CDS
Alignment Scores:
Pred. No.: 3,21e-106 Length: 52498
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
DB: 1 Gaps: 1
US-09-825-414-66 (1-487) x AF232004 (1-52498)
QY 1 Methisileasnglseralaglnp-propoglyvalalamegluserpheargthr 20
|||||
DB 46117 ATGCACATCAACGACGCGCTCCACACACCGCTGTGACTGCGACGATGACTTTGGACA 46058
QY 21 AlaseraspalaseralaseraserValargserValserThrThrserysarg 40
|||||
DB 46057 GCGTCCGACGCGCTCTTGCTCCAGCTCTGTGCGATCTGTGACGTCGATGACGACGAC 45998
```

```
QY 41 AspleglnAlaIlethrAspTyrIleuysHSHISValAlpheAlaAlaIsrpheser 60
.....
DB 45997 GAGATPAATGCGATTGCGCATTAAGTCTGACAGATCAAGTGTGCGGATTAACGCGG 45938
QY 61 ValIleGlySerProAspGluThrArgSpAlaValAlaAlaIsrAspGluGlnIleAspAla 80
||| |||
DB 45937 CCGGGCGATTGCGCGTGAATGCGCAACGTCGATGACATGACATGACATGCGGATTCACGCG 45878
QY 81 LeuValGluThrArgAlaAsnArgLeuThrSerGluGlyGluThrProAlaThrIleAla 100
|||||
DB 45877 CTGATGAGACCGCGCGACCGCGCTGACATTCGAAGGGAAACCCCGCAACCATTCGCC 45818
QY 101 GluThrPheAlaValAlaGluThrPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
.....
DB 45817 GACACCTTCCGCGCAAGCGGAAAGCTCAGCGATTTGGCGACGATACACGACGCGTTG 45758
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnIlys 140
|||||
DB 45757 CCGGGACGCGCTTGGCCATGGCCCTGCTTCAATGATACGACCTGCGATCAACGACG 45698
QY 141 GlyAspThrPheAlaThrProLeuIysProLeuThrProLeuIleSerGlyAlaLeuSer 160
|||||
DB 45697 GCGGATTGGCTGCGCGCTCCGCTCAACCGCTGACCCGCTCATTTCCGAGCGCTGTCG 45638
QY 161 GlyAlaMetAspGluValGlyThrIlysMetAlaPargAlaArgGlyAspLeuHISThr 180
|||||
DB 45637 GCGCGCATGACGACGATGGGACCAAGATGATGACCGGCGACGCGGATGTGCAATTGC 45578
QY 181 LeuSerThrSerProAspIysLeuHISAspAlaMetAlaValSerValIysArgHISser 200
|||||
DB 45577 CTGACGCGCTGCGCGGACAGGCTCCACATGGGATGGCCGCTTGGTGAAGCGCACCTG 45518
QY 201 ProAlaLeuGlyArgGlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
|||||
DB 45517 CCAGCGCTTGCCTCGACAGGCTTCTGACACAGCGGCGTTCGCTTCAACGACTCGCGCGCC 45458
QY 221 AsnValAlaArgThrValLeuAlaProAlaIleAlaSerArgProSerValGlnGlyAla 240
|||||
DB 45457 AACGCGGTGACGATGATGCTCGCGGACCTGGGCTCCAGACCGCGCGTGCAGGGGTGT 45398
QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
|||||
DB 45397 GTGGACCTTGCTGATCATGATGGCGGGTGTGCTGCGTCCACAGCAGCGCTTGGCAACGC 45338
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValIleuGlyMetIys 280
.....
DB 45337 CTGCTCAGTGTGCTGCTGCTGATCACACGCGTGGCGGTGCTTGTAGTCTGCTTGAAG 45278
QY 281 AspIysGluProIysAlaAlaLeuSerGluIuThrAspThrLeuAspAlaThrIysAla 300
|||||
DB 45277 GATTAAGAGCCCAAGGCTCACTGAGCGAAGAAACGATGCTGAGAGCTTATTAAGCA 45218
QY 301 IleIysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyIysArgMetAlaGlyLeu 320
|||||
DB 45217 ATCAATGCGCGCATCTACGCGGTGCGCGCTCAACGCTGCGAAGCGATGCGCGGTGCTG 45158
QY 321 ProLeuAspValAlaThrAspGlyLeuIysAlaValArgSerLeuValSerAlaThrSer 340
|||||
DB 45157 CCACGTGATATGCGACGACGACCAATGGGTGCGGTAAAGACCTGTGTCACGCTCCAGC 45098
QY 341 LeuThrIysAsnGlyLeuAlaLeuAlaGlyIysArgIysAlaGlyIysSerIysLeuIlys 360
|||||
DB 45097 CTGACCCCAAAAGGTCGTGGCTGCGGGGTGCTTTCGAGGGGTGAGCAAGTTTGAAGAG 45038
QY 361 MetAlaThrIysAsnIleThrAspSerAlaThrIysAlaAlaValSerGlnIleuSerAsn 380
|||||
DB 45037 ATGGCGAGGAAAAATATCACACCGCGGACCAAGGCGCGGTGATGTCAGTTGACCAAC 44978
QY 381 LeuValGlySerValGlyValAlaPheAlaGlyThrThrThrAlaGlyLeuAlaThrAspPro 400
|||||
DB 44977 CTGGCAGGTTGGCGACGCGGTTTTCGACAGGCTGAGACGACGCGCGCTTCAACACGATGCC 44918
```


Db 1174 ATCAAAATCCGCCAGCTACTCCGGTGCAGCGCTCAACGCTGGCAAGCGTGTCCGGCGTCTG 1233

Qy 321 ProleuaspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
|||||
Db 1234 CCGGTGACGTCGCCACCGATGTTGTAAGACCGGTGAGGAGTCTGGTGTGGCCACCGC 1293

Qy 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrTrpAlaGlyValSerLysLeuGlnLys 360
|||||
Db 1294 CTGATGCAAAACGGTCTGGTAATGGCGGGTGGTTTCCGGTGTGGGTAGATTGCAGGAA 1353

Qy 361 MetAlaThrLysAsnLierThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
|||||
Db 1354 ATGGCGCGAAATATATACACCGCCAGCCAGCAAGCGCACCGCTCAAGCCAACTGACCAAT 1413

Qy 381 LeuValGlySerValGlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspPro 400
|||||
Db 1414 CTTCGAGGTTTCAGCCCGGTTTTCAGCTGGACAAAGGACGAGCGGGGAGGACCCCA 1473

Qy 401 AlaValLysLysAlaGlySerPheIleGlnAspLysValLysSerThrAlaSerThr 420
|||||
Db 1474 GCGGTGAAAAGGCTGAGTCTTTCGACAGACAGCGTGAACAACGATATCCACGCGG 1533

Qy 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
|||||
Db 1534 ACCAGCGCGCTGGCGGCAAAACCGTCACAGCTGGCCAAAGCGGCGCATTCAGCGAGCGCC 1593

Qy 441 GluAlaIleSerThrArgLysAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
|||||
Db 1594 GAAAGCTTGTCCACTACCGGGGCGCTCTTGCAGACAC-----TTGCGCGCGTGC 1644

Qy 461 SerAlaProGluAlaAspIleGluGlyGlyTyrIleSerAlaPheSerArgSer----- 478
:::
Db 1645 ACGGTGCGTAGCGCTGATATCGAGAGGCGGTCTTCCCGGCGACGGGTTCGCGGTT 1704

Qy 479 GluThrProPheGlnLeuArgArg 486
|||||
Db 1705 GCAGTGCCTTTCGAGCGCAATGCGG 1728

RESULT 3
AL646083 208050 bp DNA linear BCT 07-DEC-2001
LOCUS
DEFINITION Ralstonia solanacearum GMI1000 megaplasmid, complete sequence;
segment 8/11.
ACCESSION AL646083 AL646053
VERSION AL646083.1 GI:17431588
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 208050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brotier,P., Camus,J.C., Cactolico,L.,
Chandler,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavye,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siglier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 208050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, IMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope et CNRS
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Blometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT Christian.Boucher@toulouse.inra.fr

FEATURES
source
misc_feature
gene
CDS

http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1. 208050
/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
135..947
/gene="Rsp1115"
/note="RS05505"
135..947
/gene="Rsp1115"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD18267.1"
/db_xref="GI:17431589"
/translation="MADSVYOQVADVPGLVLSAGRFQSFERRHLDHFVGLVTAG
VORRFKCKTVLIGPCISLIAMPREIHQSHVAGDQAVTKTRLSQALIANLAADISG
THREPLATLIEDRALAGHLRLHDAHQSHASATLQVSEVALALATLTKSRALIP
EHFKGLSPAQWRVNDYCFHAGERITLDDLASICALRFQFLQKQTVMTBHW
LVRLRLEQASLISRSKRAIAEVAQEVGYDQSHFRNFRQAFVAPSHYQS"
1162..2259
/gene="Rsp1116"
/note="RS05504"
1162..2259
/gene="Rsp1116"
/EC_number="4.1.2.5"
/function="small molecule metabolism; degradation; amino
acids"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Homology
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOLASE
PROTEIN"
/protein_id="CAD18267.1"
/db_xref="GI:17431590"
/translation="MNRDPMPTTRTPALGFTSDNTEGASPEVLDAIILANAGAAPYC
ADCTARVQTLSELEFEVAVFLVPTGIANALCLGARTPPWGVYTCIPSSHINDE
CGAPERVYNGAKLVAVDGPASAKIDVEKLEAAVRVGVGHSTQPCVSLTQATEVGL
VYLAEEVIALGQDKARIKRIDGDSRFANALVSLCSPAAVMKAGVDALISGATRGK
VVAEEVIEVILSLAEMGYRRKRAGLFESKRFSAQIEATLADMLRNARQANDM
ARLRLEGMEGASVLTGATEVITFCRPSPLIENLLQAGRFRFHNHREPVVAFVT
SPATRADVDNILLRMKAAAPSDTHTATA"
2257..5019
/note="ACUR or RS05503
Predicted by Framed"
/evidence=not experimental
complement(2462..3139)
/gene="Rsp1117"
/note="RS05502"
complement(2462..3139)
/gene="Rsp1117"
/function="miscellaneous; unknown"
/note="Product confidence : putative
Gene name confidence : hypothetical
Predicted by Codon_usage
Predicted by Framed"
/codon_start=1
/evidence=not experimental
/transl_table=11

gene
complement(3188. .3595)
/gene="Rsp1118"
/note="RS05501"
complement(3188. .3595)
/gene="Rsp1118"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18269.1"
/db_xref="GI:17431592"
/translation="MEDTKIAIVVRBDLATWOKNTAFIMSGTASPEIMGAYRD
ADNHYSLICQVPVIVLSGDQETLKAHRALEKRVPHGLYIEDMFSTGHDNAARAVF
AEHGPD SARVVGIALREKKIVDKITTKGARMP"
complement(3673. .4161)
/gene="Rsp1119"
/note="RS05500"
complement(3673. .4161)
/gene="Rsp1119"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Homology
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18270.1"
/db_xref="GI:17431593"
/translation="MGMSGEVEVCPIARTLAVVGDRTVLLIRELPLGVKREEFGA
OTGMSHLSTPRKLEADGIVVRHLYSDRPREHYRLTSGDFYPLLSTLKWGK
WGFQFAKTAALITTHRCGHGTGLKLVCPACDDEPFGKDKITVLGASFAAERQARD
GA"
4228. .4725
/gene="Rsp1120"
/note="RS05499"
4228. .4725
/gene="Rsp1120"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18271.1"
/db_xref="GI:17431594"
/translation="MPALTRHPRCGAPRYLEDKVVGLDTETSGFAVTRDMILSFAEY
DQPMHIDEIARDVFCGLVSGWOTLAVYMRLLDRLILGTPIVGAEILRIDIRFIA
PKRPEDVIRAREVIAIRPSKSRPDRGMDARVTTNSEGTLVTIGMSLVVPRPDE
OKVAF"
complement(5017. .5610)
/gene="Rsp1121"
/note="RS05498"

gene
complement(3188. .3595)
/gene="Rsp1118"
/note="RS05501"
complement(3188. .3595)
/gene="Rsp1118"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18268.1"
/db_xref="GI:17431591"
/translation="MNDASLHETIIISCFLEHORPPVREINAEERNRDEAOYRQALRAL
ADNHSLICQVPVIVLSGDQETLKAHRALEKRVPHGLYIEDMFSTGHDNAARAVF
AEHGPD SARVVGIALREKKIVDKITTKGARMP"
complement(3673. .4161)
/gene="Rsp1119"
/note="RS05500"
complement(3673. .4161)
/gene="Rsp1119"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Homology
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18269.1"
/db_xref="GI:17431592"
/translation="MEDTKIAIVVRBDLATWOKNTAFIMSGTASPEIMGAYRD
ADNHYSLICQVPVIVLSGDQETLKAHRALEKRVPHGLYIEDMFSTGHDNAARAVF
AEHGPD SARVVGIALREKKIVDKITTKGARMP"
complement(3673. .4161)
/gene="Rsp1119"
/note="RS05500"
complement(3673. .4161)
/gene="Rsp1119"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Homology
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18270.1"
/db_xref="GI:17431593"
/translation="MGMSGEVEVCPIARTLAVVGDRTVLLIRELPLGVKREEFGA
OTGMSHLSTPRKLEADGIVVRHLYSDRPREHYRLTSGDFYPLLSTLKWGK
WGFQFAKTAALITTHRCGHGTGLKLVCPACDDEPFGKDKITVLGASFAAERQARD
GA"
4228. .4725
/gene="Rsp1120"
/note="RS05499"
4228. .4725
/gene="Rsp1120"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence: hypothetical
Gene name confidence: hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD18271.1"
/db_xref="GI:17431594"
/translation="MPALTRHPRCGAPRYLEDKVVGLDTETSGFAVTRDMILSFAEY
DQPMHIDEIARDVFCGLVSGWOTLAVYMRLLDRLILGTPIVGAEILRIDIRFIA
PKRPEDVIRAREVIAIRPSKSRPDRGMDARVTTNSEGTLVTIGMSLVVPRPDE
OKVAF"
complement(5017. .5610)
/gene="Rsp1121"
/note="RS05498"

```
Db 204151 GTGCGCCCGCTGGCAACCCATGCCCTGGGGCGGCAAGGCGAGCAAC-----GTGGAT 204204
QY 243 Phg1yValSerThrAlaGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeu 262
Db 204205 TCCTGGATCCCGCGGTGGCGGGCCCGCGGTGGCGCCGCTACATAGCCATGCGACG 204264
QY 263 SerValGlnSerArgAspGlnLeuArgGlyValAlaPheValLeuGlyMetLysAspLys 282
Db 204265 CACATGAAATAGACCCACCAACCCGACCGGCGGCGAATACCTGCTGGCC----- 204312
QY 283 GluProLysAlaAlaLeuSerGluGluThrAspThrPheLysAspAlaThrLysAlaLeuLys 302
Db 204313 -----CCGACCGACTGGGAAGACACAGTTCAAGCAGACTCAAG 204348
QY 303 SerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu 322
Db 204349 CAATCGACGTGACGTGCCCGCTGTGGGAGCGGCGCAACCGCACCGCAAGCTGCTGCTC 204408
QY 323 ---AspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db 204409 GCGGATTTGGCGACCGGACCGGACGCTGGCCGACGCTGTTCACCGCCACCAACCTC 204468
QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyLysArgMetAlaGlyLysMet 361
Db 204469 ATCAAGAACTGGGCGGCGACGTGGGCTTTGCCGCTGTGTGACCGCACAGACCGCT 204528
QY 362 AlaThrLysAsnLeuThrAspSerAla---ThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 204529 GCGGCGAAGCCCGCCACGAGGCGCGCTATACGAGCGCGCTCCGCCCGCCGACGCGC 204588
QY 381 LeuValGlySerVal-----GlyValPheAlaGlyThrPheThrAlaGlyLeuAla 397
Db 204589 GCGCTCGACCGGTACTGTGCGACCGGCTGACGCTGCGTGAGCAACGCGCATGTGATG 204648
QY 398 ThrAspProAlaValLysLysLysAlaGluSerPheIleGln 410
Db 204649 GCGGCGCCCGCATCGACGCGCGCGGCGGCGGCGACATCCAG 204687

RESULT 4
LOCUS AY094894/c 1558 bp mRNA linear INV 15-APR-2002
DEFINITION Drosophila melanogaster R67967 full insert cDNA.
ACCESSION AY094894
VERSION AY094894.1 GI:20151774
KEYWORDS FLI.CDNA.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE 1 (bases 1 to 1558)
AUTHORS Scapleton,M., Brockslein,P., Hong,L., Aghayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dreanek,D., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarini,H., Krommiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Patel,S., Phouanavong,S., Wan,K., Yu,C., Lewis,S.E., Rudin,G.M.
and Celniker,S.
TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
COMMENT Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720
This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may
```

have not been detected in our initial analyses such as internal priming, priming from contaminating genomic DNA, retained introns due to reverse transcription of unspliced precursor RNAs, and reverse transcriptase errors that result in single base changes. For further information about this sequence, including its location and relationship to other sequences, please visit our Web site (<http://fruitfly.berkeley.edu>) or send email to cdna@fruitfly.berkeley.edu.

FEATURES

source

1..1558
Location/Qualifiers

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db_xref="taxon:7227"

/map="64C5-64C5"

1..1558

/gene="CG13705"

/note="alignment with genomic scaffold AE003482"

/db_xref="FLYBASE:FBgn0035582"

59..1357

/gene="CG13705"

/note="Longest ORF"

/codon_start=1

/product="R67967p"

/protein_id="AAM11247.1"

/db_xref="GI:20151775"

/db_xref="FLYBASE:FBgn0035582"

483 c 404 g 373 t

BASE COUNT 298 a 483 c 404 g 373 t

ORIGIN

Alignment Scores:

Pred. No.: 0.185 Length: 1558

Score: 160.50 Matches: 97

Percent Similarity: 38.52% Conservative: 59

Best Local Similarity: 23.95% Mismatches: 157

Query Match: 6.67% Indels: 92

DB: 3 Gaps: 15

US-09-825-414-66 (1-487) x AY094894 (1-1558)

QY 141 G1yAspThrPLeuAlaThrProLeuLysProLeuThrProLeuIleSerG1yAlaLeuSer 160

Db 1363 GCGCGCTTACGCGGACGAGCTTCAGACAGCAGCGGCTTTGACGCGATCGCT 1304

QY 161 G1yAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis--- 179

Db 1303 GCGCAGACGCGTGTGCTGCGCGGCGACATCAATGATGGCAGAGCGGCTCTCCACGG 1244

QY 180 -----TylLeuSerThrSerProAspLysLeu 188

Db 1243 GCGCAGCTACTCCGCGAAGCAGCTCAGACATCAGCGCGCGGCGATCCACAGCAGC 1184

QY 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGly 204

Db 1183 CACGGCTTGTACGCGATCATCATCAGCAGCGAGGAGTGTCCACAGCAACAGCTCGGA 1124

QY 205 ArgGlnValValAspMetGlyIleAlaValGln-----ThrPheSer 218

Db 1123 GCGGACCTCTCAACAGGCGGCGATATTCAGACAGCTCGAGACATCAGCAGCAGC 1064

QY 219 AlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238

Db 1063 ACGGATCACGACGACGCGCTTTGTAGCGGTATTCATCTACGACGAGAGGATCTCGAC 1004

QY 239 G1yAlaValAspPheGlyValSerThrAlaGlyLysLeuValAlaAsnAlaGlyPheGly 258

Db 1003 GGC-----AACGGCTTCGGA 989
 QY 259 AsparMetLeuSerValGlnSerArGaspGlnLeuArgLYLYAlaPhaVallleuGly 278
 Db 988 GCGCGATTCTC-----AACAGGGGCGAGGATTTTCGGCGGACAGCTGTTAC--- 941
 QY 279 MetLysAspLysGlnProLysAlaAlaLeuSerGlnGluThrAspTrpLeu----- 295
 Db 940 -----ATCAGCAGCGGAGGATTCACACAGCAGCGAGCTTGTAGCGGTA 896
 QY 296 -----AspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
 Db 895 TCCATCAGCAGCAGGAGGGGGTCTCAACACAGCCTCGAGGAGAGCTCTCTCAACGGG 836
 QY 314 GLYLSArgMetAlaGlyLeuProLeu---AspValAlaThrAspGlyLeuLysAlaVal 332
 Db 835 GGGCAGGATCTCGCCGGAAGAGAGCTCATTCACATCAGCAGCAGCAGATCAGCGGCG 776
 QY 333 ArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu-----AlaLeuAla 349
 Db 775 GACGGCTTGTAAAGGTACATCATCATCAGCAAGAGGGGTGTCACAGCAAGAGCTCAGA 716
 QY 350 GlyGlyTyrAlaGlyAlaSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
 Db 715 GCGGAGCTCTCAACAGGGGGGAGGATTCAGAGGGAGCCTCGAACAATCAGCAGCAG 656
 QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyAlaPheAla 389
 Db 655 ACGGATCACACGGCGGACAGCTCTGTAGCGGTAGCATCATCATCAGCAGGAGAGCTCGAC 596
 QY 390 GLYTPTrpThr---AlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPhe 408
 Db 595 GGCACAGCGCTCGAGGAGGAGCTCTCAACAGGGGCGAGCTACTCAGGAGAGAGTTC--- 539
 QY 409 IleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaLaspGlnThr 428
 Db 538 -----GGAGACATCATCAGCAGCAGCAGCAGCAGCAGCAGCTCTT 497
 QY 429 ValLysLeuAlaLysThrValLysAspMetSer-----GlyGlnAlaIleLeu 444
 Db 496 GTAGCGGTGATCATCATCAGCAGGAGGAGTCTCGACAGCAGCAGCTCGAGGAGGAGCTC 437
 QY 444 rSerThrGlyAla-----SerLeuArgSerThrAlaAsnAsnLeuArgHisArg----- 460
 Db 436 CTCACAGAGGGGAGGAGTACTCAGGAGGAGTTCGAGACATCATCAGCAGCAGCAGTAC 377
 QY 461 -----SerAla 462
 Db 376 ACGGCGGAGGCTTGTAAAGGTACCATCATCAGCAGCAGGAGTCTCGACGGCAAGCGC 317
 QY 462 apProGuaLaspIleGluGluGlyLysSerAlaPheSerArgSerGluThrProPh 482
 Db 316 TTCGGAGGGGATCTCTCAACAGGGGCG-----AGTATTCACCGAGAGGAGAGCTCTT 263
 QY 482 eGlnLeuArgArg 486
 Db 262 CACATCAGCAGCG 250
 RESULT 5
 ACT108877/c 163709 bp DNA linear INV 22-JUN-2002
 LOCUS Drosophila melanogaster 3L BAC RP98-17L24 (Roswell Park Cancer
 DEFINITION Institute Drosophila BAC library) complete sequence.
 ACCESSION ACT108877
 VERSION ACT108877.2 GI:21538999
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 163709)
 AUTHORS Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Gocayne,J.D., Tabor,P., Williamson,A., Homsli,F.H.,
 Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
 Ayele,M., Scott,G.S., Morley,K.W., Amanatides,P.G., Braddon,R.C.,
 Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Bunay,C.,
 Busam,D.A., Center,A., Chen,G., Chen,Z., Clerehugh,B.,
 Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.B.,
 Draper,H., Emery-Cohen,A., Ferreira,S., Garg,N.D.S., Houck,J.,
 Hostin,D., Howland,T.J., Hume,J., Idegami,C., Jaitli,M., Kovar,C.,
 Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
 Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
 Scheele,F., Shen,H., Strong,R., Tector,C., Wang,O., Williams,S.M.,
 Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
 Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.
 Direct Submission
 Unpublished
 2 (bases 1 to 163709)

TITLE
JOURNAL
REFERENCE
AUTHORS

Worley,K.C., Adams,C., Adio-Oduola,B., Alt-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbata,J.,
 Benton,J., Blmage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Franco,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Harl,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
 Hollins,B., Homsli,F., Howard,S., Huber,J., Hutik,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korah,J.,
 Kovar,C., Kralovic,J., Kuresh,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtenage,O., Lieu,C., Liu,J., Liu,W.,
 Loulseged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,
 Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
 Mel,G., Metzker,S., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S.,
 Oguh,M., Okwuonwu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
 Ruiz,S., Savery,C., Scherer,S., Scott,G., Shen,H., Shoshitari,N.,
 Sisson,I., Sodergren,E., Sonake,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,Y., Villalob,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wiczysk,R., Wooden,S.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G. and Gibbs,R.
 Direct Submission
 Unpublished
 3 (bases 1 to 163709)

TITLE
JOURNAL
REFERENCE
AUTHORS

BCM-HGSC.
 Direct Submission
 Submitted (01-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 4 (bases 1 to 163709)

TITLE
JOURNAL
REFERENCE
AUTHORS

BCM-HGSC.
 Direct Submission
 Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jun 22, 2002 this sequence version replaced gi:18463997.
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
 gc-help@bcm.tmc.edu

COMMENT

segment 3/11.
AL646078 AL646053
VERSION
AL646078.1 GI:17430778
KEYWORDS
ORGANISM
Ralstonia solanacearum.
Ralstonia solanacearum.
Bacteria; Proteobacteria; beta subdivision; Ralstonia group; Ralstonia.
REFERENCE
1 (bases 1 to 203050)
Sainoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S., Ariat,M., Billault,A., Brottier,P., Camus,T.C., Cattolico,L., Chandelier,M., Choisme,N., Claudel-Renard,C., Cunnac,S., Demange,N., Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T., Siuier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M., Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 203050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie Molculaire des Interactions Plantes-Microorganismes INRA-CNRS, BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS 118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27, F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire INRA, BP27, F31326 Castanet-Tolosan Cedex
COMMENT
Christian.Bouche@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
1. 203050
Location/Qualifiers
/organism="Ralstonia solanacearum"
/strain="GM1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(147..1499)
/gene="RSP0310"
/note="RS05460"
complement(147..1499)
/gene="RSP0310"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not-experimental
/transl_table=11
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CAD17461.1"
/db_xref="GI:17430778"
/translation="MNAARKNTDRAEALALFNASTHEPANPCWIPDAAAAGLHPGRWRARFMAIFGGALSLIGSALTOFLWMITDTGSVALTAGIAALLPDAVISPGLGTFADRSRLIMTIVADGISALCMILLIALFLTRIALMHAYMMAIRSAQAQAFQAAASAVTMVPGGFLVRAAGLNOSLSLIVAAIPGALAIQVMPIGMAGIDVATPAGIAPLICRIPOACVSSGOKTGLMPAREGVDAIWKPKRMLTVLLGAVVLTJMPFTLVPLVKTTHFGGAAOVALMEGSLGIGMAGVAGLLVAMARRROYANTLIGFANSCVALATPALAGSLGVAVAMWVTSITTFVGNAPLTLQTTIVPNHLOGVPSLMTWMLAAPVGLAFLTEVIGVRAIFAIGLGAASLAGLSPSLMKLRDRPSEARND"

predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not-experimental
/transl_table=11
/product="PROBABLE OUTER MEMBRANE CHANNEL LIPOPROTEIN"
/protein_id="CAD17462.1"
/db_xref="GI:17430780"
/translation="MRLIPLIVALTASCSLPPALIKPALPVPATPPAESGPEFOAHAADIGWRPFQDRRLQRLIGTLINNRDLRLAALNVERAYOYSTORARRLPGIDATGSASRRRTANDGSAALAAASEQVGVGVGATPEIDFGFRSLSDPAFARYLSDOGKRAQISLVGAVADATFAERLADQSRLOAHTLADHRQSIDLARKLAKONSDLYVAQAEQVMAADADLEARTRALAQGNSIQRLVGGAGLPADLPAPLSLETQPVVTLDPALPSEBILIRRPDIRQAEHMLAANDDIGARAPPRSTLTASLGVASPMGSLPDGRRVMSFPQITPDLFOGRLRAELRLAEKRSASIAEYERTIOFAFEVADGLARATYGROIJEORAVASAREVELSNLRACVEERLELDSQRLVYASQALDLPQGAFSNVAALYKALGGGLAETVDATVADRVARK"
complement(2966..6115)
/gene="mexD"
/note="RSP0312; RS05457"
complement(2966..6115)
/gene="mexD"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not-experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE PROTEIN"
/protein_id="CAD17463.1"
/db_xref="GI:17430781"
/translation="MPQFTIRPVFANVILFTLLGALAIKPLRTARYSVAPPTYNLSATVPATQTMNDGVISLIERLSSVKNLTFRESSDTSASATVATKPGCDPM AOVDVQNKLAIEBRLPOVYKONGLIVESASGFLMTIGQSDGRDELDALDYPM NVBELRIEIGVGVOLEFGAEQAMRIWDEPAKLIAYLSKNDLSAALSOONAOIAPBS VGASPALPAORVTVLFGQOLTPEQEFALIVRANANGSKVMGDAVRELGAOSYG PINENGRKATPAAVOLAPGAMNVRTGEVQARBAELSRPMGMSYSTPNNAPFK ISTEKVOTILEANVLFVLYLFLQVNRVTLIPALYAPIALICTFTIMLASGSESTIV LTFMGVLAIGITVDDAIVVENENVERIMAOEGLSPRATIKAKETGAVIGITLVLT AVFLPMALAGSVGVIIQOFTLSMAVSIVESAFALITLPALCATLTKPDPAHHEKR GVFEFMRRDRITOGIYAAVNGKLVYRTGVMVFGAIAAVLALAEFROLPSSELPED OGYEMTFEOLPADTAERTLDDVYKAFERHVASRASISQNSIYVFGSGSSNAAMF TMLKMDNRGATATNONEMRLAQAAMAGATGTMVSPMIDELGTSSTGTLRQDA NQGDALKAETKLELAOSRIYGVYPPGLPAGTISRIDIDRQKALGAVPTSTIS ETTSTAIGSLYVNDFFNAGRMQOVTIADAPARKQIDVILKTVRNAGGAVPLSEVY RPVWSETPQLQVREGPPARISGNAPASGSAAMEMERLAAQLPAGPAILMTQGS LOEROSAAQAPMLAALSMLVYFLVLALESMAIPLSVMLVVPGLGAVLAVLLRMGM PNDVFFVGMITVYIGLSAKNALILVEFKALIDREGKGLVESAVTAARLRLRPILMTSL AFGIGVYPLMIAMGASAEFOHALITGVFGWMSATVLAIFEVVFEFVVGIDBERFEA WRASROAPPPPVYQROEG"

343 sasnglyleualaleualagly----- 350
Oy
Db 52306 CAGCGCGCGCTGCTGTGGGCTTCCGGGCGAGCGTGACCGTACCGCTACCGGCAACGCCAGCGC 52305
Oy 351 ---glytyralaglyvalserlyslauglnlysmetala----- 362
Db 52366 CACCACCTACAGCGGCGCGCCCTATACGACGAGGGGGGAGCGTATTCGTTCACGCGCAT 52425
Oy 363 -----7h 363
Db 52426 CCAGATGCTGCTGACCGCGCGCGCCCAACGACACCTTCACCGCTCCGCCAACAC 52485
Oy 363 rlysaenlhetraspsersalathrlysalalalavalserglnuserasnlheuval-- 382
Db 52466 GGCCCAACGACGAGCGGCGCAACAGCCAGCGCTTCGGCGAGCTGCGCAACGCCACGCT 52545
Oy 383 -----glyservalglyvalphealaglytrpthrthralaglyleualathras 399
Db 52546 GCTCGACAACGCGCACCTCGATCGCTGGGCTGAGCAACCTGCTGACCCAGGTGG 52605
Oy 399 ptyralalavallyslasagluserphelleglnasplysvallysserthralaserse 419
Db 52606 CATCCAGCGCACCGCGCGCGCAACCTGACGTCGCAAGAGCGCGCTGCGCAAGTTTC 52665
Oy 419 rthrtsertryvalalalaspplnthrvallysleu--alalythrvallyaspm 438
Db 52666 CACGTCGACGACGACGCTGCTGCGGCGCTCACTGACGACGACGACGACGACGACGACG 52725
Oy 438 tserglytluallalasersestrthglyalaserleuargserthrvallasnleu 457
Db 52726 GAAGTACACGACGCGCTACGACGCGCTCGCGCAAGGTGATGACAGCGGCGACCTCGCTG 52783
CDS

RESULT 7
AE008061/c 10029 bp DNA linear BCT 18-DEC-2001
DEFINITION Agrobacterium tumefaciens str. C58 circular chromosome, section 119
of 254 of the complete sequence.
ACCESSION AE008061 AE007869
VERSION AE008061.1 GI:15156405
KEYWORDS
SOURCE Agrobacterium tumefaciens str. C58 (Cereon).
ORGANISM Agrobacterium tumefaciens str. C58 (Cereon).
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
1 (bases 1 to 10029)
REFERENCE Hinkle, G., Slater, S.C. and Goodner, B.
AUTHORS Complete Genome Sequence of Agrobacterium tumefaciens C58
TITLE (Rhizobium radiobacter C58), the Causative Agent of Crown Gall
Disease in Plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10029)
AUTHORS Hinkle, G., Slater, S.C. and Goodner, B.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2001) Cereon Genomics, 45 Sidney Street,
Cambridge, MA 02139, USA
FEATURES
source Location/Qualifiers
1..10029
/organism="Agrobacterium tumefaciens str. C58 (Cereon)"
/strain="C58"
/db_xref="taxon:181661"
141..1142
/gene="AGR_C_2485"
141..1142
/gene="AGR_C_2485"
141..1142
/note="hypothesized protein - Synechocystis sp. (strain
PCC 6803)"
/codon_start=1
/transl_table=1
/product="AGR_C_2485p"
/protein_id="AAK87138.1"
/db_xref="GI:15156406"
/translation="MLADKGCDAGRDRAVCCSSRNDEQETDVALNXYDVFTEAKM

AGNPLAVYADADLSQDTMQLAREMNLSEVFTRNSTNSHAASRLIFPSPGELPRA
GHPVGAVALAIAERNRGDODDIDMVCVGVPAVCAYVMRAFAPFADLPKRS
SRVPLPNTALADALGVSECHGHEENHVPSTIMTAVGPFLVPHNIIAAYDMFEDN
LMUTAPLVESELAAATYICRGVNHAKPHARFSPBEMGAEDBPATGSAALASGLT
HHFDGLTDGHTPLLIEQGVEMERSPHILRMDIKDNEIARIRIGHAAVRAVASTFEL"
1364..3313
/gene="AGR_C_2489"
1364..3313
/gene="AGR_C_2489"
/note="OXYTETRACYCLINE RESISTANCE PROTEIN"
/codon_start=1
/transl_table=1
/product="AGR_C_2489p"
/protein_id="AAK87139.1"
/db_xref="GI:15156407"
/translation="MRTNLGSLIAHVADGKSLTERLLTFQVGYIDKIGSVDTGWTQD
SLELERQGITIRAAVSVFTIGDVTNLLIDTPGHPDIAEVERYLGILDAVYVVSAY
EGVOAQTRVLRALQRLAVPLPFTIMVDRVARYDEVLRDLAQRLVRPVMSKTTG
AGSKHVOAATDTECEPLFTILCTLENDDELBDYLLPDRDAGRSLILGQRTA
CGVHPAPAGVAMTAGLIPALIAATEMLPARDPDEGEISGKLFKIERGGGCKLSY
LHLSGTYRLRHIIPLPGPARLIGTOLFPRGRVONANSISLAGOIAHYTGLAGARVGD
SVGDVANADGAFHFRAPTLERVRYSRPSSEAALMLANQVREDPLIGLRNDETNE
VFSLYGEVQKQITQSELSTGFIIEFEDSTVYICVERLSSGGLOTIFEEPFPFA
TIGLREVRPDGANSFALIEVEFGQMPASEFYRAVEVFTDLKQGVFGWQVDHVM
TAARHTSPSSTAADPRKLTLPWLIGTALQADPFVCEPFDHFIHAPAAITRLISLA
KAGATKNSVTSIGGVATLEGTIASAAVQGVQOVPGLTSGIGCMETSFSSHYAOESP
SPRRSGDPFNECDYLLRMRNAE"
complement(3406..9873)
/gene="AGR_C_2490"
complement(3406..9873)
/gene="AGR_C_2490"
/note="2490"
/codon_start=1
/transl_table=1
/product="AGR_C_2490p"
/protein_id="AAK87140.1"
/db_xref="GI:15156408"
/translation="MAELISLEGFRMHGDSPLADRKVMGMRPGVCTVRNRMAN
KISDVDETAFOALEDALQALFAFEKPEPTRKAPKQPEARVKAASQTPAPEATRP
VEQTPRSPNLEAANDGSKRSPAMLTSLGEGSIGGALRNATIMSVINAGLIGLTAHL
YGNALMSGSLADLTATPGTMAITVGLTVPMLPFAPIAMARADLRNARSAAEVA
LRLAEPETVASDRITMSVGOAVRREVSAMNDICTETIATATELETVHSEVALRSTA
DNELRVRLVQELTAEKDAIVNHAERIKRSITVGAQEOIKELSTVGEELSMRIATGGE
AFASIMIDPRSAALLKESRASTAEAGSLIAATEMLALQVNSGSIISNEPFRMLNLT
STLBERGVLLERPAIHASTDISGVEISNLSERTROLNLTLSARSLERLNRETCO
QVIGSGLDVTVDLSTLLEKGLSFRSOSIGTADATIDMDLSIEGEEERQAAVVOG
NSAPDEHYAQAASAPDORAGSLDKIMSLARINIEYVAGSGEADITLTSLERIGST
MTDOSLALATALGTQEMLENALSRITQESDAIGQRTAETITDPAFTNSHAKITDYLAE
RSMALFGALSQDFEDALASRLATIGSVGAHEHLAALIDERRAALINSVADVER
RLTETLETRAAATIGAVSGIEMDIPTLEERTALAHVSGAERISADITLDGRTAALS
SAISGVEERADTMDSTRISLDMTFANEEESTLMDRTSALTGIYASAEKTAGAL
DSRTATGDEVVAGAETRIETLDGRTAALNPDVSGAERIDALPDVSGAEDRTMDLSRTMA
EERKAEADPTTALGELVAAETRINGALNPDVSGAERIDALPDVSGAEDRTMDLSRTMA
LDMSPSGVEEKITDILDGRTAALKSAVGVDRTAGALDSTRTALSGVSGAERITAE
ALDSTRTALDMTISGVEERIAEAMDAARSLSJLAAVGQRLTAFTLLENALASGHE
RLFTMLGQAEIRIGSLEENSGLEIOSVGAANRINENVVEDGSSRFQTVBEGVSRLE
NNLSOSHEIRTALDROADRLAATLSATTOGMDSLOAMMIGTIVYASSMSMLSL
ETQODTLOKATIDGSAITLRLRNSAGDIYKIGEAAREIGAGATSTRITSETISGV
TTRIDETCARLETSLDALQTVRGCDIANVNSJIDAGRFPALDEDTAVAPARNSDEA
AERTIGLIDETTVADTFENRISRLAETPAGTARIDERTGMDRALTGLEVNRT
IEGRASDLAVSLRGAVVATONIGDEARSSALAKSGSEAEVQVQONETPAIEE
RSGEIVTRISDAQTRLGSOAAVQTFSEAGNIIVNVYAEAEVAVRSQVGISETLRS
VESALDARGESIRSLADNRTRELNSMLASRAEELRIIEKAKPVBEVATIGREAAE
KIVAAOOSAEILSOSNGWYGVJEQIISQVAAAGTAAASLVAATROSTOMLSETH
AMADAVEDGIDTKYRSGSDANKUYAATRODMLSOTHNMAITVEDSATNNAVAY
RAAGVFGADALNASTRFESASQADWSSSRLBLEGTIDRLSINSQGLTAVQAG
IGVRFEEHNSKYLQASBELNLAQSLVGTGFERDRLASLTSVGLVKSBELETAMRY
IGVVEENTLNEAERSQNVAGNLNDNLASFEDIRDSLETSPORARSAQTRGALLSA
GODASRSIESTLSDAOKXSDLVNLRKGVSSISSEVNDLGSASEKSNAAANIKET
LROAVEEVSREACATDEIRRSSHDIRELDITAEELKRGAGLPEEKKEESAAMRA
VSEQIKALODISQVGRSTHOMEVSEPARAIATOPRAEERVYPRPQAPAAAVOOR
PAOPAPALRGTLPLENROVENROADPAOPAVATNPGREEGGWISDLIRASOE
TPAASTPRASTEDOPTRAADTRNRRHVESJNSVDIARIDHDASVEIMRRVROGE

RDVETRLTYLKGGTTTDEIKRKYREAEFRFAVDRYITIDEKLLADVARTRDRSVT
OSYLSDDGKVTYMLAHAGREN"
BASE COUNT 1772 a 2913 c 3262 g 2082 t
ORIGIN

Alignment Scores:
Pred. No.: 10.8 Length: 10029
Score: 149.50 Matches: 96
Percent Similarity: 34.70% Conservative: 65
Best Local Similarity: 20.69% Mismatches: 176
Query Match: 6.21% Indels: 127
DB: 1 Gaps: 16

US-09-825-414-66 (1-487) x AE008061 (1-10029)

QY 69 ASPALALALEUALAHISANGLNIIIEASPAALALEUVALGLUTHRARGALASARG 88
DB 8049 GATGCCCTTCACCAATTCGACGCGAAGATGATACCGTTCTGGCGGACGACGACGCGC 7990
QY 89 LEUTYRSERGLUGLUTHRPROALATHRIIEALGLUTHRPHEALALYSALAGLUYS 108
DB 7989 CTGTTTC-----GGCGCGCTTTCGGCCAGCCAGACGCGT 7957
QY 109 PHEASPARGLLEUALATHRTHRILASERSERALAPHE-----GLUASNTHRPRO 124
DB 7956 TTCGACGACGACACTCGCGCGCGTCTCTCGCCATTACGGCGCTTCGCGACGCGC 7897
QY 125 PHEALALALASERVALLEUGLINTYRMETGLNPROALALISANLYS----- 140
DB 7896 GAACACCTTTCGCGCATCTCATGACCGCGCGCGCATCAACAGCGTCGTTGCCGAC 7837
QY 141 ---GLYASPTRLLEUALATHRPROLEUYSPROLEUTHRPROLEUILESERGLYALALEU 159
DB 7836 GTTCGAACGTCGCTGACGAGACGCTCGAACCAGCCGCGACGCCATCAACGCGCGTTC 7777
QY 160 SERGLYALAMETASPOLINVALGLYTHRLYSMETMETASPARGLALARGLYASPLEUHS 179
DB 7776 TCGCGCATGAGAGACCGCATCTCCGACGCTGGAAGCCGACGCGTCC----- 7726
QY 180 TYRLEUSERTHRSERPROASPLYSLEUHSASPALAMETALVALSERVALYLSARGHIS 199
DB 7725 -----CTGCATGATGTCGTTTCGCGCGCGGAGAGACCGCATC 7690
QY 200 SERPROALALEUGLYARGLINVALVALASPMETGLYILALVALGLNTHRPHESERALA 219
DB 7689 GCGGATACGCTTGAC-----GGCCGACCGCGCGCTTTCCACGCGG 7648
QY 220 LEUASNVALVALARGTHRVALLEUALAPROALALEUALASERARG----- 234
DB 7647 ATTTCGGGTGTCGAGAACGTAATTGCCGACGATGACGACGCCGCTGTCACATCGAC 7588
QY 235 -----PROSERVALINGLYALVALASPPHNEGLYVALSERTHR 247
DB 7587 ATGACCTTCGCCAATGTTGAGAACGCGCTTCGAGACGCTCGACAAACGCACTCCGCT 7528
QY 248 ALAGLYGLYLEUVALALASN----- 254
DB 7527 CTGACAGGGATGTTGCAAGCCGCCGAAGAAAGATCGCGGCGCTCGATAGCCGACG 7468
QY 255 ALAGLYPHEGLYASPARGMETLEUSERVALGINSERARG-----ASPLINLEUARGLY 272
DB 7467 GGCACCTTCGGTGGAGCGTGTTCGCGTCTGAACCCGTAATGCCCAAAACGCTTATGGC 7408
QY 273 -----GLYALAPHEVALLEUGLYMETLYSASPLYSGLINPROLYSALAALA 287
DB 7407 CGCACCGCAGCGCTCAATCCGCTGGTTCGCGGAGAGAGCGC----- 7363
QY 288 LEUSERGLUGLUTHRASPTRLPLEUASPAALATYRFLYSALAILLEYSERVALSERIYSER 307
DB 7362 -----ATGCGGATCTCTCGACGACCGCTACCATGCGCTCGACATGACCTTCTCC 7312
QY 308 GLYALA----- 309

DB 7311 GCGCGCCAGGAAAGATCGCCGACGCGCTCGACACGCGACCGACCGCTGTGAGCTG 7252
QY 310 ALALEUASNALAGLYLSARGMETALAGLYLEUPROLEUASPAALATHRASPGLYLEU 329
DB 7251 GTGGCAAGCGCCGAGACCGCTATTCGCCGCG-----GCCCTCGACAGCCGACGATTCGCTG 7195
QY 330 LYSALAVALARGLSERLEUVALSERALATHRSERLEUTHRYSANGLYLEUVALALEUALA 349
DB 7194 AAAACCGTC----- 7186
QY 350 GLYGLYTYRALAGLYVALSERLYSLEUGLNYLSMETALATHRYSANGLIETHRASP 369
DB 7185 -----GTTCCGCGCGCGAAGAGCGGATCACCC---GATGTCGTCGACACG 7144
QY 370 ALATHRYSALALALAVASERGLNLEUSERASNLEUVALIGLYSERVALGLYPHEAL 389
DB 7143 CGCACCATGCGCTCGACATGAGCTTCTCGCGCTCGAGAAAGATCAACCGATATTC 7084
QY 390 GLYTRPTHRTHRALAGLYLEUALATHRASPPOALVALLYLSALAGLUSERPHEIL 409
DB 7083 GATGCCCGCACCGCC-----GCCCTGAAGAGCCCGCTTCCGCGTGT 7042
QY 410 GINASPLYVALYSERTHRILASERSERTHRTHRSERYVALALASPOLNTHRVAL 429
DB 7041 GAGGACCGCATTCGCGCTCGACACCGCTACGTCGCCCTTCGCGC----- 6991
QY 430 LYSLEUALALYSTRVALYLSASPMETSERGLYALALILESERSERTHRGLYALASER 449
DB 6990 ---ATCGTGTCAAGCGCGAAGAACGACATCGCGCGCGCTGATGACCGCACCTTGCA 6934
QY 450 LEUARGSERTHRVALASNLEUARGHISARGSERALAPROGLUALASPLIEGLUGLU 469
DB 6933 CTCGACATGACATYTTCCGGTGTGCAAGAGCGGATCCGAGCGATGATCGAGAGCC 6874
QY 470 GLYGLYILESER 473
DB 6873 TCGTCGCTTCG 6862

RESULT 8
AE009096/c 12020 bp DNA linear BCT 20-DEC-2001
LOCUS
Agrobacterium tumefaciens strain C58 circular chromosome, section
DEFINITION
122 of 256 of the complete sequence.
ACCESSION
AE009096 AE006688
VERSION
AE009096.1 GI:17739754
KEYWORDS
Agrobacterium tumefaciens str. C58 (U. Washington).
SOURCE
Agrobacterium tumefaciens str. C58 (U. Washington).
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

REFERENCE
AUTHORS
1 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., KitaJima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boyee Sr.,D., Chapman,P., Glendenning,J., Deatherage,G.,
Gillet W., Grant,C., Guenther,D., Kutysvin,T., Levy,R., Li,M.,
McClelland,E., Palmeri,A., Raymond,C., Rouse,G.,
Saenphlammachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,J., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
TITLE
The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
JOURNAL
Science 294 (5550), 2317-2323 (2001)
MEDLINE
21608550
PUBMED
11743193
REFERENCE
2 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., KitaJima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boyee Sr.,D., Chapman,P., Glendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutysvin,T., Levy,R., Li,M.,
McClelland,E., Palmeri,A., Raymond,C., Rouse,G.,
Saenphlammachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,

FEATURES
source
1. 12020
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
68. 538
/gene="Atu1343"
68. 538
/gene="Atu1343"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA142349.1"
/db_xref="GI:17739755"
/translation="MGQALYIVDQPSFNPASLVAEISALAGTMPISATVERHDESY
TPERQLGMRPGRDVSLVPADRFIFKRGYPPREVIDHLSTLKDRLVLCGIGADTC
VLAAGFALPDAGLHPTLLPMLTVGSSLSDRSGELCARLKHFGAVLGPALAG"
complement(566. 1066)
/gene="Atu1344"
complement(566. 1066)
/gene="Atu1344"
/note="putative; ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA142350.1"
/db_xref="GI:17739756"
/translation="MHVKTLLAALLATTVPAPARQTAIDTPVSAIEVSSGME
DRNLEPVKGDGQLRPPAPSPTRGKYVIAIROGDGAKIKLRIVFTADPSLENT
ELEEFSGMKSYVTVDVREPSNGASASGLFTVYTLKIDPAKKAWSMTILDELGEKK
VERASN"
complement(1086. 1859)
/gene="Atu1345"
complement(1086. 1859)
/gene="Atu1345"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA142351.1"
/db_xref="GI:17739757"
/translation="MSHSTYSLSFVASPEBAOTAKLEKRAVYSDTPPEADYVALG
GDFMLQILNETMNSGKRVYGNMGVSGLFNDNRVGLLQRIKIVASGNDHPLRMTT
TDSGDDEFTALANNEVSLFRQSHQAALRVEAGLREELICGMVAVAPAGSTAVN
FSAFGPIPLPESPLALTLPVSAFPRRMGALIPNKKTVDLHVLEBDKRPVNAVADHT
EKSQVHRTIAOSODRTAKILSDPRSDRVALEQFN"
2164. 3081
/gene="Atu1346"
2164. 3081
/gene="Atu1346"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="phenazine antibiotic biosynthesis related
protein"
/protein_id="AA142352.1"
/db_xref="GI:17739758"
/translation="MALNYDIDYDFTAKKAGNPLAVNYDADISQDTMOAIAEMNL
SEYVFNKSTNPISHAASLRIETPSGELFPAGHPVGAVALAERNREGDDIDMVCY
LEEKVGVRCAYKMRAGAVAFEDLPKSSRVLPINHAALADALGVSEGLGFENH

VPSTWAGVPELLVPMHNIAAVRMDMDANLWLTATVEGLLAAYIYRGGVNHA
KFAHMFSPEDAIEDPATCSAAALGAIHHEFDGLDGHYPLIEGVEMERSHIT
LRMDIKNEIARARIGHAVAVASGTEEL"
3303. 5252
/gene="Atu1347"
3303. 5252
/gene="Atu1347"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="tetracycline resistance protein, tetM/teto
subfamily"
/protein_id="AA142353.1"
/db_xref="GI:17739759"
/translation="MRTNLGILAHVDAGKTSLERLLPVGVIDKLSVDTGNTOTD
STLEERGITIRAAVSTFTIGDVVNLIDTPGPDIAEYERVLGLDAAYVVSAY
EGVDAQTRVLRALQRLAVPPLFTINKVDVGARDVDRDLAQQLRVRPVSKITG
AGSKHGVAAITDECEBLFTILCTELANDDELRLDLPDRVDARSLLLQOQTA
CGVHPAFAGVAMTAGLIPALIAITMPLRADPDPEIGISGKIFKIERMGGEKLSY
LHLASGVRLRHILPLPOGPARLGIOLFQKGRVQNSIGAGQIAHVTGLAGARVD
SVGVDAVADGAFHAPPTLETRAVYSRPSSEALMLNMGVEODPLIGLRNETNE
VFVSLYGEVOKQIIQSLSNGFGEAEFEDSTVCYRLSGSGGLOFTIRREPPFA
TIGLRVPERDGAENSRFALEVEFGQMPASTRYAVEETVDTLKGVGKMOYODCHAM
TAARTSPSSTAADFRLTPWLTGTLAKQAPFVCEPFDHFLIAPAAATRLISILA
KGAATNSVYISGCVATLEGTIASAAVQGVQVQVGLTSLGGMETSPSHYAOESP
SPRRSGPDPFNECDYTLRMRNAE"
complement(5345. 11812)
/gene="Atu1348"
complement(5345. 11812)
/gene="Atu1348"
/note="Identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA142354.1"
/db_xref="GI:17739760"
/translation="MABLLSLEGKFRMHGDSPLADKKVNGMKPGCVTRRNMANN
KISDVETAEALDELALOGAFEEKEPRTAKQPEKRVKASQTPAPETRAP
VEOPRSPNLEAANDSGKSPAMTKLEGGISGALRNATIMSWAIGLIGTJHLL
YGNLMSIGSLADTALPGLAIYGLIYVPMLEFPAIIMARRDRDNRNARSABEVA
LRLAEPETVSDRIMSVAQVAREVSAMNDIETATKATELETLVSEVNALESTTA
DNELRVSLVOELTAERDAIVNAHERIRSSVQAOEDIKELSTVGELEMLRIATTE
AFASMIDTRSAALLEKSRASIEAMGSLIAKTEMLLOALNMSGSTIENEDMRJLNT
STLDERGEVLLERPAIHASTLDSGVEEUNLEERTLOMNETSARSELERNTERGO
CVTGSIDTVYDKLSTLEKGLSFRSLOSTADADIMDLIRGLYEERMOAVGVQ
NSARDENVAQFASAFDORAGSLDSKLMESLARIETVYAGSGEALDTLITSGLEIGST
MTDQSLALATRLGTOQEMLENALESRRQAFSDAIGQRTAETTDATFTNSHAKIDTVLAE
RSNMLFALSASODRFDEBALASRLIATGVSAGAEHLAAMLEBAAAINSVADAVR
RLUTTELETPRAAITGAVSGIEDRISDTLESRTAAIHDVVGAEBSRIADTLDRTAALS
SAISGVEERTADTMDSRTLSIDMTFANVEERTLIDNRRISALGUYAASAEKJAGAL
DSRTATFGDVAVAGAEIRLAIETLIDGRTALAAVYSGAERTIRADLSMTALDGEASA
EERKAEALDRTALRGLVLSAERTIRAGALDSRDSIKTYVSSGAEERTITVDLSRTMA
LDMSFSGVEERTITDIDGRTALKASVAGVBDRLAALDSRTAALSIVSGAEBSRIAE
ALDSRTALVEMITJSGVEERTIRAEADDAVASISLAAVGGORLETAFTLINALSGHE
RLFETMLSSQERIASLERNSGLEIDGVSQAANRIENVVADGSSREFQVTEEGVSRLE
NNLSOSHETRLTALDORADLAATLSATGOMGMLEQOAMIGTIVAASSMSEL
ETQODTQKALIDGSAATLEARLRNSADIAVKIGEARREIGANDALSTRERSIGV
TTRIDETGARTESLDALQTRVGGDLANVNSISDAGRRRADLEDTAYFARTSDRA
AERTITGILDQTRVADTFENTRSRLAETDAGARIDERGTDRALTTGLEVNNT
IEGRKADLAVSLRGAVVSATONIGDEARSAALIAKSGSEBAQVQONAEFTKAIIE
RSGEIVTRISDAOTRSGOAAVAVQTESEAGNIIVNVAEVAVRQVQVYISITLNS
VESALDARGESIRBALDNRRLNLSMIASSASAEIRLIEKARPVVEVATIGEEAE
KYSAAOOSAEILLSQNSGAWGAYEQALISDIAAGCTAASKVAIATQSDMLSETIT
AMADAVDQIGDKYARSGSDAANKLVATROSTDLSTHNSMEIVQDSATNFNAAVE
RAAQGFADAEALNASTATRESASQADAVSSSRLIEKIDRLSKISSEETLAAVNG
IVGRFEESKVLISOASELLNAASQSLGLEERODALRSLSGVLYKRSSEETLAAVNG
VGVVENTLNEAEERSONVAGNLBNLADISGSDISDETDEORARSAAOMRGLLSA
GODASRISTESTLSAQKYSDELVYRLKGVSESSIDEPNDILKGSSEKSNAAANLKT
LRQAVEEVSFPAGATDEIRSSHDIRELDATAEKLRGAFDLPEAKSAAAMARA
VSDQIKALODISOLVGRSTHOMEVSEVARIATQPAARRVPRPPAPAAAVOOR
PAQOPAPALRGTLPLBNRQVENROAPAPAPVATINAGRREGGCIISDLRGAQOE

BASE COUNT 2153 a 3504 c 3864 g 2499 t
ORIGIN
TTPASTRASTEEOPTRAAATNRNHRVESINSISVIAIADHDAVEELMRROEE
RDAFTBRLLYTLKGTOTFDEELKKRYERAEERTIANDRIITDEKLLADVARDRDSVT
QSILTSPTGKAYITMLAHAGRN"

Alignment Scores:	
Pred. No.:	13.7
Length:	12020
Percent Similarity:	149.50
Matches:	95
Conservative:	66
Mismatches:	176
Best Local Similarity:	20.098
Indels:	127
Query Match:	6.218
DB:	1
Gaps:	16

US-09-825-414-66 (1-487) x AE009096 (1-12020)

Qy	69	AspAlaIaLeuAlaHisAsnGluInIleAspAlaLeuValGluThrArgAlaHisnArg	88
Db	9988	GATCCCTTCACCAATTGGCCAGCGGAAGATGATACCGTTCTTCGGCAGACGACGACGCC	9929
Qy	89	LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys	108
Db	9928	CTGTTC-----GGCGCGCTTGGGCCAGCCAGGACCGT	9896
Qy	109	PheAspArgLeuAlaThrThrAlaSerSerAlaPhe-----GluAsnThrPro	124
Db	9895	TTCCACCAAGCACTCGCCACCGCCCTTCTCCGCAATACCGGCTCGGTTTCGGCAGCGGC	9836
Qy	125	PheAlaIaIaSerValIleuGlnTyrMetGluProAlaIlaAsnLys-----	140
Db	9835	GAACACCTTCCCGCATGCTCATGAACCGCCCGCGCATACAGCGCTTCGCCAC	9776
Qy	141	---GlyAspThrPheAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeu	159
Db	9775	GTCGACAGTCCGCTGACGAGACGCTTCGAAACCGCGCGCAGCATACCGCGCTGTC	9716
Qy	160	SerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis	179
Db	9715	TCCGGCATCGAGACCGCATCTCCGACACGCTGGAAAGCCGACGCGTGC-----	9665
Qy	180	TyrIleuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis	199
Db	9664	-----CTGCATGATGTCGTTCTCCGGGGCGGAGACCGCATC	9629
Qy	200	SerProAlaLeuGlyArgGlnValValaIAspMetGlyIleAlaValGlnThrPheSerAla	219
Db	9628	GCCGATACGGTTAC-----GGCGCGACCGCGCTCTTTCAGGCGC	9587
Qy	220	LeuAsnValAlaArgThrValIleuAlaProAlaLeuAlaSerArg-----	234
Db	9586	ATTTCGCGTGTACGAGACGTAATTCGCCGACAGATGACAGCCGACGCTGCACCTGCAC	9527
Qy	235	-----ProSerValGlnGlyAlaValaIAspPheGlyValSerThr	247
Db	9526	ATGACCTTCCCATGTTCGAGAACGCGTTTCGAGACCTCGACACCGCACCTCCGCT	9467
Qy	248	AlaGlyLysLeuValAlaAsn-----	254
Db	9466	CTGACAGGCAATTTGTGCAAGCGCCGAAGAAAGATGCCGGGCGCTCGATACCCGACG	9407
Qy	255	AlaGlyPheGlyAspArgMetLeuSerValGlnSerArg-----AspGlnLeuArgGly	272
Db	9406	GCCACCTTCGGTACCGGGTTCGGCGGTGTGTAAGAACCGATGCCGGAACGCTTATATGC	9347
Qy	273	-----GlyAlaPheValIleuGlyMetLysAspLysGluProLysAlaIa	287
Db	9346	CGCAGCGACGCTCAATGCGCGGTTCGGGTGCCGAGAGACGCG-----	9302
Qy	288	LeuSerGluGluThrAspThrPheLeuAspAlaTyrLysAlaIleLysSerAlaSerThrSer	307
Db	9301	-----ATCCGCGATGCTCTCGACACCGTACCATGGCGCTGCACATGACCTTCTCC	9251
Qy	308	GlyAla-----	309

Db	9230	GGCCCGCAGGAAAAGATCGCCGGAAGCCCTCGACACGCGACCCGACCCCTGTGTAGCTG	9191
Qy	310	AlaLeuasnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu :::	329
Db	9190	GTGGGAAGCCCGAGACCCGATATGCCGGC---CGCTCGACAGCCGCACGGATTCGGTG	9134
Qy	330	LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 	349
Db	9133	AAAACCGTC-----	9125
Qy	350	GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 	369
Db	9124	-----GTTTCGGCGCGCGGAAGAGCGGATCC--GATGTCTCGACAC	9083
Qy	370	AlaThrLysAlaAlaValSerGlnLysSerAsnLeuValGlySerValGlyAlaPheAla 	389
Db	9082	CGCACCATGGCGTCGACATGAGCTTCGCGGGGTGAGGAAAAATCATCCGATATTC	9023
Qy	390	GlyTyrPheThrAlaGlyLeuAlaThrAspProAlaLysLysAlaGluSerPheIle 	409
Db	9022	GATGGCGCGACCGC-----GCGGTGAAGAGCCCGTTCGCGGTGT	8981
Qy	410	GlnAspLysValLysSerThrAlaSerSerThrTyrSerTyrAlaAlaAspGlnThrVal :::	429
Db	8960	GAGGACCGCATTCGCGGTGCGCTCGACACCGCTACTGCTCCCTTCGCG-----	8930
Qy	430	LysLeuAlaLysThrValLysAspMetSerGlyGlnAlaIleSerSerThrGlyAlaSer :::	449
Db	8929	--ATCGTGTCAGCGCGGGAAGAACGATCGCGAGCGCTGATAGCGCACCTTGGA	8873
Qy	450	LeuArgSerThrValAsnAsnLeuAlaArgHisArgSerAlaProGlnAlaAspIleGlu 	469
Db	8872	CTCGACATGACAAATTCGCGGTGCGAGACGATCGCCGAGCGATGATCGAGAGCC	8813
Qy	470	GlyGlyIleSer 473 :::	
Db	8812	TCGTCGCTTCG 8801	
RESULT 9			
AE008285		12313 bp DNA linear BCT 18-DEC-2001	
LOCUS		Agrobacterium tumefaciens str. C58 linear chromosome, section 89 of	
DEFINITION		187 of the complete sequence.	
ACCESSION		AE008285 AE007870	
VERSION		AE008285.1 GI:15159303	
KEYWORDS			
SOURCE			
ORGANISM		Agrobacterium tumefaciens str. C58 (Cereon).	
REFERENCE		Agrobacterium tumefaciens str. C58 (Cereon).	
AUTHORS		Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;	
TITLE		Rhizobiaceae; Rhizobium.	
REFERENCE		1 (bases 1 to 12313)	
AUTHORS		Hinkle,G., Slater,S.C. and Goodner,B.	
TITLE		Complete Genome Sequence of Agrobacterium tumefaciens C58	
REFERENCE		(Rhizobium radiobacter C58), the Causative Agent of Crown Gall	
REFERENCE		Disease in Plants	
AUTHORS		Unpublished	
TITLE		2 (bases 1 to 12313)	
REFERENCE		Hinkle,G., Slater,S.C. and Goodner,B.	
AUTHORS		Direct Submission	
TITLE		Submitted (14-AUG-2001) Bioinformatics, Cereon Genomics, 45 Sidney	
JOURNAL		Street, Cambridge, MA 02139, USA	
COMMENT		Approximately 800 bp of telomeric sequence missing from the left	
COMMENT		end of the chromosome and 200 bp missing from the right end.	
FEATURES			
SOURCE		Location/Qualifiers	
gene		1..12313	
gene		/organism="Agrobacterium tumefaciens str. C58 (Cereon)"	
gene		/strain="C58"	
gene		/db_xref="taxon:181661"	
gene		113..895	
gene		/gene="AGR_L_1737"	
gene		113..895	
gene		/gene="AGR_L_1737"	

/note="PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C
REDUCTASE)"
/codon_start=1
/transl_table=11
/product="AGR_L_1737p"
/protein_id="AAK89441.1"
/db_xref="GI:15159304"
/translation="MSLPKIGFIGTGATIDAMVRGLLVKPAAPVPHVMYSQSRADISA
RLAADPQVITSRNOMIVGCDVMIAIPQIAEYIRLPRDQKIVSYAATSR
DALIDMGAVRLAOLIPLEVARRKGVTAIPADPTIAIFNVIGNAVCTTKAEYD
LIAASALMATYRGIMOTTEMLAENLPEPDKARAVYPLFAGLSFVALLSGKTDPT
EMSRFRATKGLNEQVMQDFRNGSGNALKEALSREWERTAQ"
complement(951..2549)
/gene="AGR_L_1740p"
complement(951..2549)
/note="probable MFS transporter PA2701 (imported) -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_1740p"
/protein_id="AAK89442.1"
/db_xref="GI:15159305"
/translation="MSAKSSASSFAPLAQPVPAVYMAATVLTNGTSEMDVASSWIL
TDLASPAVALVQAAGTLPFLAIPAGVLTIDRKKFLIAVOLLSVSTLAWL
AHTGMLSVSALIGLTFGIGALMGPTQOAIYDELKREDIKSAVLANLSYTLAWL
IGPAVGIIILAAFGAATYGADVAYEFVIAALIMPRKANNAALOENFEGFRAGL
RYTRASRLHVILLRAAIFPAFASAWALLPLVARLTLGGDASFGTILGAVGATA
GALVMTKRRAPDDCLLGAATVYTAIVMAGLSFAPPKYATITLLPLGAVTALUT
LNGAQAIIIPNVYRGGLAVLTIVNCAATAGSISGWAVEAGAVPOTLILGAGLET
AGLIMHRLKLPAGDADNVPSNHMPEPLVAEPVAHDRPVLILLEYTEKHRSAPFLA
LIDLSQERRRDGAVGWGVTEDSADPOKIVEMFVSWAEHLRQHKRVSNADADLOGKV
LAYHSGLEKPVYHFLITINRGA"
complement(2546..2986)
/gene="AGR_L_1741p"
complement(2546..2986)
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_1741p"
/protein_id="AAK89443.1"
/db_xref="GI:15159306"
/translation="MTDPSSTLRGLGLAEIVAPLRTVALALACSAIYQGPLTKIF
DFPGALAEHMHFGLOPAAPAEVFIIEPLTASAVVYGLVRWGAALASGFTLLAPFI
ALRPWEMAPGMDMMATNAFPEHGLAGATVYIAIDLTGKGTGK"
complement(3196..5073)
/gene="AGR_L_1745p"
complement(3196..5073)
/note="hypothetical protein PA2699 (imported) -
Pseudomonas aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_1745p"
/protein_id="AAK89444.1"
/db_xref="GI:15159307"
/translation="MHDDTLHNGRVTLDRGNPNATATAIKDGLFEVSGDAVVAL
AGSETKIIDKGRVLPGLIDNTHYIRGSLINRMELRMDCVSLADAMMLKROVAL
TPAOWRVVGGFTTEHOFVEKRLPTEIEINAAVADPTVFLHLIDRALGALRAAG
YTKDTPMPGGEITRDANGPRTGLLAKPNAAGILYSTLAGPKLPIDLOYONSRHEFR
ELNRLGTVGVIDAGGFQNPDDYEVLOKLSDEQOMTVRLAYNLFTQKPOEEDFN
WTSVVKYKQNDYFRHNGAGMLVEFSADPEDFPOAPDMPPEMEGELDEVVRLAEN
RMPMRLLHATYDETISRALDVEFKVYNRDIPLEGLNWEPDVAETISDRSIDRLAAGGI
ATQGRMAVQGEYFVERGCGVAEATPTPIAMLEKGVASGTDATRYASINPNVSLFM
MVTGKTIVGGMLTPRANCLDRETLARMTKRVTFWSSEBKGKTEKGQDPLVPEPK
DYFSCADEISLFTSELTMVGKIVYGADGFRVLDENINPPAMPDMSPVPTFGGYAM
GEPOGAGCKNSLRRAITSCGASCAGVGHHDHAGAMTSKLPIDLKGFGALGCSGWA
V"
complement(5178..5531)
/gene="AGR_L_1747p"
complement(5178..5531)
/gene="AGR_L_1747p"

/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_1747p"
/protein_id="AAK89445.1"
/db_xref="GI:15159308"
/translation="MPRRKQYPVRRGFIHPKRVSKVYLLSLGAGLVGYISLLNV
RSPAPPVIALVGLGILVGEQIIPFARSIIIDREPAVSWFNQIKPMFGHMPGKAET
TDLAQOQOSNRKS"
complement(5542..6192)
/gene="AGR_L_1749p"
complement(5542..6192)
/note="probable hydrolase PA2698 (imported) - Pseudomonas
aeruginosa (strain PAO1)"
/codon_start=1
/transl_table=11
/product="AGR_L_1749p"
/protein_id="AAK89446.1"
/db_xref="GI:15159309"
/translation="MIFIDQOPQMARFVOSIDROTLLKNVYGLAKAAKIFNPTITTT
VFNESFGNFPPELLAVPENDILERTSNMSMDQDVRAALAKNAAGTRKIVGSLM
TEYCNTPFALSADHDVADYETIVVADASGCTSIDAKRYAMDRVQAGVPIYMOOVL
EMORDARKETVAVTSLVKEHSGAIGMDIAVTNVHGAERYKAGKRLGPWPAK"
complement(6370..7305)
/gene="AGR_L_1752p"
complement(6370..7305)
/note="ABC transporter homolog ydfo - Bacillus subtilis"
/codon_start=1
/transl_table=11
/product="AGR_L_1752p"
/protein_id="AAK89447.1"
/db_xref="GI:15159310"
/translation="AMIKDKIKGLHHTSNASDARQNNSEPTDTLGRVRYKTYNDEP
SVYHLTYGDEMGTPGVMTTFPPDHMAAGRGVGEVGFOTFSTPKAMGFWKRLTAM
GAAGIADIVFGANRLRFTGPPDGLALLETANDGRAWQAEIISEMNLKGTSGARF
NLNDTATEIELLRFNGYERADAGNVTYRVIPGNGADITDLATLNTPTFARGACSV
HNLFAFVNERKOLEYRKALMDGYOTVPIYDRYFMAIYFRTPGILFEIATNERGE
DRDDTAHGEALKLPSRYEPFRNQIEANVSLAA"
7421..8320
/gene="AGR_L_1754p"
7421..8320
/gene="AGR_L_1754p"
/note="(AF007569) Gstr"
/codon_start=1
/transl_table=11
/product="AGR_L_1754p"
/protein_id="AAK89448.1"
/db_xref="GI:15159311"
/translation="MNDHKALKIFLMAAEKRNFAOVARLSMTAGVTRAIALAECDL
GVLQFVTPRVOSLITDDGAFEAQIOIPAYEALNSOVARDRNAKADGGRISAPWIL
GKTVPPLIISGRFELPKKCFEISLSDGLVNIYDDYDLAIRSSAPSKFTWKRIR
AVPRILVAAGSRFTQMKHPSDLTPDDCLAYSESRRENNITLSDGSSITTAGRAFS
ANNGEVLADMAADAGVAMLPGFHIAEHRSGRLVHLPQMAPPELMLTLYPPYOTL
PPRIATFSKFEFEEDLTAHMYALD"
9016..12045
/gene="AGR_L_1758p"
9016..12045
/note="1758"
/codon_start=1
/transl_table=11
/product="AGR_L_1758p"
/protein_id="AAK89449.1"
/db_xref="GI:15159312"
/translation="MGSWQOQIVTIRHSVQTFMTSLSSISAVOISKLSTSTIAGL
TSEDVVALSTKIKALSSQIASLSTDNALFESEDLRAISYAVVGLPLTQWAKLSS
AOYSSLSQVATLYASOIDAMSTOIKALNSOVAGLSSAOVATLSDBELALFTTDE
IKSISANATAGSAAAIAGLSTONAAATRKSOIATAMSSTOFNALTSGSLATFSADVK
AISNKTILAGIDYTKLSTGNIRALSKAQVSLSTTQRAASTDQIKALTSTQOVAGLSSA
QVATLSSDELATFDEIKRIGANAAVAGSAAALVALTTDDNAALTKTQIAGLSTYOL
NALTSLATFSADEIKRISTKALAGLDVTKLSTGNIAALTKTQOASTLSTQFAAMST

VERSIONS	KEYWORDS
AEO009328.1	GI:17742413
SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington).
ORGANISM	Agrobacterium tumefaciens str. C58 (U. Washington) Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Rhizobium.
REFERENCE	1 (bases 1 to 12668)
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitaajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Boye, S.D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, M., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphumachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V., and Nester, E.W.
TITLE	The genome of the natural genetic engineer Agrobacterium tumefaciens C58
JOURNAL	Science 294 (5550), 2317-2323 (2001)
MEDLINE	21608550
PUBMED	11743193
REFERENCE	2 (bases 1 to 12668)
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitaajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Boye, S.D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, M., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphumachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V., and Nester, E.W.
TITLE	Direct Submission
JOURNAL	Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA
FEATURES	Location/Qualifiers
Source	1..12668
gene	/organism="Agrobacterium tumefaciens str. C58 (U. Washington)" /strain="C58" /db_xref="taxon:18085" complement(85..5280) /gene="ina" /note="Atu3974" complement(85..5280) /gene="ina"
CDS	/note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer" /codon_start=1 /transl_table=1 /product="ice nucleation protein homolog" /protein_id="AA144776.1" /db_xref="GI:17742414"
	/translation="WTSITLALNNOIRYISTEAVVAAMTEDVASLSTQIKALSSDQ IALDVEDVKILNSQQLSAISOGAIVGLTDDQNIILQYAKRSLSASQVSLTTPQF ALTDDQAEVLTSSQVRVYLSAQALASGDDIATSTADMAITKAMPGSGTEVIACI TPDQIALINSAISALFTTEQIALISLTDQIEALSTSOVAALSPQVRLSTGQALSTG DIAAIGKATIGTSTIEAIELTSDRAVAALSAIGITGVLSLALSTAOIALITTAQ IALKTHVALSTSOVEALSPLOYRSLPASQVLALESDEIATSTADIALITTAQ RLSTEAIGTLTLDIALITTAQALNSASQQAISAGQVLESTHIALAGVLSIASL TADQIEASTROVEALSAQVYKRLNSQALISLSPDIATFSTADIALIAGVLSIASL GAILAFTTAQIAVETGAIALISLTDQIEALSTSOVAALSPQVRLSTGQALSTG LATPSADIALITGKAMGLNTVDIALITTAQIAFQSAIAGITSPQVLEALSSQVSL LSTQIKALSTSOVAALGSVVALSTSOQVLAISLKGITLTPQIALRPDQSALE TAQIAALSSGQITAMGASVNLASLSMAQIAALISTSSMGLSTGQIAALSTGEI QMAALSSQIAALIGTEAVNALSTAOIALISGLSAAGISGKSGQVATSTAOEALSTADQ LPSFMSHIAELGTEDLAKFTTKDIAIASSISAIGLSAIALSTTAQIAALNTQISINAA LSTQIAALITTAQVLEALSTQVNALSTQKALSTQVLAITDIAITPSTQDIAINSDAIAIGLS AETIASATGQIILANVSTALSTSTQVATTAQVVALTTVOGALSTSTQVLAISLTD DIATFTSDMAALSSAIGLSQTVASLTGAIGLSAAGISGLSTGQIALTGDDQ NLVLTNTQIAALTSKQVAAPVSDITTAALSTQIILASGAAGITLTDIALSTQVGA

gene
complement(5496. .6653)
/gene="glf"
/note="Atu3975"
complement(5496. .6653)
/gene="glf"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="UDP-galactopyranose mutase"
/protein_id="AAL44777.1"
/db_xref="GI:17742415"
/translation="WADRIATVYGAGISVITRELAAGHLVDITDARDHAGNCH TERDESEVMVAVHPHIFHTDDREVMWYVNSFQTEPMYKRVATTSGEQVSLPVL HTINQFEKNEPDEARFIEDKADKSTDPQTEDEQALREVGMDLEAFKFGKQMGCSPTPEASTILKRLPVRFVNDYVDFPHKXQGMPEGSDTMEIRLIDHNITVYLGDT RDRDAEATGAGVHSPGLDGFDPDEGRLAVALRLDFEPRFDGSDYOGCAVMYNGVYS VPFRTITGKHKHSWPEHDAGSVCYRREPREGGPDIDYVPIRYLDEKDEQADYVARAE REASTVEGRGLDTHYRLDMQATIRALDTAVALYLRNSEGGSMPALHSPY"
/gene="Atu3976"
complement(6755. .8800)
/gene="Atu3976"
complement(6755. .8800)
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL44748.1"
/db_xref="GI:17742416"
/translation="WTANHVSPSNDAAAPSLADIARNORSLMDLSVSEALNAGQO AAAALKYTAFAFNNDNPTIHLAYFNVAVMNLGDRAGAAQARACLKANPEFAPQH INIAGLDELDAGLIAGVQWMSHYAEVTNVADEITGRHMTLQMGVLEBAGKTEBAQH EIALQAELEPDELPEAGOHMASLRORCKKPVLAAPSNHSLPKMDMSLITSCYA DDPMQALKAAYRLSKTILGARPDLSRPRSPKQSGGQRLRVGYSLDRLDAVVEF ALCEVLELHDKDSLEVFAYYCGNVRGDSIOBERIKAAVHCNRDIGHVDATAASQIIA DEIDLILVNGTCKDRAKIFRAYRPAPVYVFCGPDSQMSGPFHOYILISGIMLIPENK ELYSEKVALRACDQPLDRKRPVAPRAVDGGLPEDAFYVASPENGOKTTECFAMW MIILSETPGSLIMLTLTGDDVDNORLRLDAESGVALEDAFYAPAMPNDPIARIGLGD FLDFPEPGASHASTADAITSGLPVLYTNGKFFAPFGSGIYTAGVCEMISCEDEVY ARAIAGEEDROSLIYVRESIARORETSVLRDIPALARLELPMQMGCEDEGTPVPV DISNLDIYEGAGLEMLNIEFDEDSQREYRIKKLQWMDIYAPIPIIDRLMONPDN"
complement(9292. .12261)
/gene="ina"
/note="Atu3977"
complement(9292. .12261)
/gene="ina"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ice nucleation protein homolog"
/protein_id="AAL44779.1"
/db_xref="GI:17742417"
/translation="WTSITSLSAVOISKISTSTIAGLTSEDVNALDSTRIKALISSQ IASLTSDNALFSSDEDLRAISVFAKGLTLOAKLSSAQVSLSSQVYALVASQID AMSTOIKALNLSQVAGLSSAQVATLSSDELALPTPEISISNALIAGLSAAIADL SDNMAALTKSOIAMSSTQFNALTSGLARFSADDEVKAKSNLKIADIDVTKLSTGVI IAAVNSVSLSTQFPAAMSTDQIKALTSQVDAVGLSSQVATLSSDELALFSTDEIKAI IGANNVAGLSAAALALTTDNNAAALTKIQIAGVAGLSSQVATLSSNLATLFPADFKALIS

TKALAGDVTKLTGNTGNIATLTKTOAASLSSTOFAAMSTDOIKAITSEOVAGLSAQA
TUSDELAFSTDEISAFSANAAGVAGTSTALALATGNTATLTKTOIATLSSTONAL
TSGSIATSEADEKATSPKALAGLDVTKSTONVALASAOVSALTSTOFAAMSTDOI
KALTSBOVAGLSAQAATISDELAFSTDEIKAGANVAGLSAALATLTDNSA
LTKTOIAGMSSTOINATLSANATFSADIKAITKALGLDVTKSTONIALTKAO
VYLSSTOFAALSTDOIKAITSEOVAGLSAQAATISDELAFSTDEIKSTRAANA
GSAALALATLTDNSALTSTOIASLSSTOLNATLSANATFSADIKAITKALGL
DVTKSTONVALSKRAOVALTSTOFAAMSTDOIKAITSDOYAGLSAQAATISDEL
AFSTDEIKATIGATAVSGLSAALATLTDNSALTSTOIASLSSTOLNATLSANAT
FTADEKATITAKALGLDVTKSTONIALTKTOIATLSSTONVALASAOVSALTSTO
VAGLSAQAATISDELAFSTDEIKAGANVAGLSAQAATISDELAFSTDOIKAITSEO
VANDAOVEAVITAKTV"

BASE COUNT 2342 a 3472 c 4171 g 2683 t
ORIGIN

Alignment Scores:
Pred. No.: 15.8 Length: 12668
Score: 149.00 Matches: 127
Percent Similarity: 35.92% Conservative: 77
Best Local Similarity: 22.36% Mismatches: 228
Query Match: 6.19% Indels: 136
DB: 1 Gaps: 25

US-09-825-414-66 (1-487) x AE009328 (1-12668)

OY 7 AlagInIaProProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
DB 11970 GCCAGCCAGATCGACGCCATGTCGACCGACCAATCAAGCGCTCTCTCAAGTC 11911
OY 27 AlaSerSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIle--- 45
DB 11910 GCAGGCGTCAGTCGCCGCCAGTGCAGACCTGACAGCAGCAACTGCTCTTCCACC 11851
OY 46 ThrAspPylLeuIysHis-----HisValPheAlaIleHisArgPheSerValIle 62
DB 11850 ACCGACGAGATCAAGTCGATCAGCGCCCAATGCGCATGCTTCCGGGCGGCTATC 11791
OY 63 GlySerProAspGluArgAspAlaIleValHisAsnGluGlnIleAspAlaLeuVal 82
DB 11790 GCGGCTCTCGACCGACCATGACGCCGCTGACGAAAGTCTGAGTCCGCCCATCTCT 11731
OY 83 GluThrArgAlaAsnArgLeuTyrSerGluGlyIleThrProAlaThrIleAlaGluThr 102
DB 11730 TCCACGCGATTCACACCGCGTACGTCGCGGAGCTTGCACATCTTCCGGGAGCAAGTC 11671
OY 103 PheAlaIleValIleGluIysPhe-----AspArgLeuAlaThrThrAla 116
DB 11670 AAGCGCATCGCAACAAGATCTCGCTGGCTGATGTCACCAAACTCACACCGCGCAC 11611
OY 117 SerSerAlaPheGluAsnThrProPheAlaIleAlaSerValLeuGlnTyrMetGlnPro 136
DB 11610 ATGCGACGCCCTCGACAAAGGAGGTTTCGGGCTGTGCGACACAGTTCGGCTCCATG 11551
OY 137 AlaIleAsnIysGlyAspThrLeuAlaThrProLeuIysProLeuThrProLeuIleSer 156
DB 11550 TCGACCGACAG-----ATCAAGGCCCTCTACCTCCGATCCAGTCCAGTCC 11512
OY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrIysMetLeuMetAspArgAlaArgGly 176
DB 11511 GCCGCTCTCGATCCGCG-----CAGGTTGCGAGCGCTCGACGAGCGAC----- 11470
OY 177 AspLeuHisTyrLeuSerThrSerProAspIleHisAspAlaIleValAlaSerVal 196
DB 11469 GAATCGGCTCTCTTTCACCGATGAGATCAAGCGCATGCGGCGCAAGCGCTTCCG-- 11413
OY 197 LysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAla----- 213
DB 11412 GCGCTCTCTCGCGACGCGCTGGGGCGGCTTACACACCGCAACGCTGACGACATGACGAAG 11353
OY 214 -----ValGln 215
DB 11352 ACCCAGATCGGCGGCTCTGTCTTCGACACAGCTGAATGCACTGACCTCCGCCAACCTTGCA 11293

OY 216 ThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgPro 235
DB 11292 ACCTTCTCTCA---GATGAGATCAAGGCCATGACGACCAAGCGCTGGCC----- 11245
OY 236 SerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyIleLeuValAla----- 253
DB 11244 -----GCTCGCATGTTACCAAGCTCAGCAGCGGCAACATCGCGCGCTGAC 11197
OY 254 -----AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeu 270
DB 11196 AAAACCGAGCGCGCATCGCTGCTCCACACAGTTCGACAGCTGTCGACCGCAGATC 11137
OY 271 ArgGly-----GlyAlaPheValLeuGlyMetLeuAspIysGluProIlyAspAlaLeu 288
DB 11136 AAGGCCCTCACATCCGACAGGTGCGCGGTCTCACTCCGCGGAGTTGGAGCGCTCAGC 11077
OY 289 SerGluGlu-----ThrAspThrLeuAspAlaTyrIysAlaIleIysSer 303
DB 11076 AGCGACGACATGCGCTCTCTTCCACCGATGATGATCAGCGCTTACGCGCAATGCGGTC 11017
OY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAla----- 313
DB 11016 GCTGGCTTTCACACAGCGCTGCGCTGCCCTCACACCGGCAATGCAACCGCACTGACG 10957
OY 314 GlyLysArgMetAlaGlyLeu-----ProLeuAspValAlaThrAspGlyLeuIys 330
DB 10956 AAGACACAGATCGCTGCTCTGCTTGCACACAGCTCAATCCCTGATATGACGACATC 10897
OY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrIysAsnGlyLeuAlaLeuAlaGly 350
DB 10896 GCCACTTTCGCCCGCATGAGTCAAGGCCATCAGCACAAG-----GCATGCGCGGC 10843
OY 351 GlyTyrAlaGlyValSerIysLeuGlnIysMetAlaThrIysAsnIleThrAspSerAla 370
DB 10842 -----CTCGATGTGACCAAGCTCAGCAGCGGCAAGCTGCGA---GCCCTC 10801
OY 371 ThrIysAlaIleValSerGlnIleSerAsnLeuValGlySerValGlyAlaPheAlaGly 390
DB 10800 AGCAAGGACAGATTCGCGGCTGCGACGACA-----CAGTTCGCTGCC 10756
OY 391 TrpThrThr-----AlaGlyLeuAlaThrAsp 399
DB 10755 ATGTCGACGACGACAGATCAAGGCCCTGACATCCGACGAGGTGCGCGCTCAGCTCCGC 10696
OY 400 ProAlaValIysIysAlaGluSer-----PheIleIleAspIysValIysSer 415
DB 10695 CAAGTTCGACGCTCAGCAGCAGCAAGTGGCTCTCTTCCACCGATGATGATCAAGCGC 10636
OY 416 -----ThrAlaSerSerThrThrSerTyrValAlaAsp 426
DB 10635 ATCGGCGCCACCGCGTGGCGGCTGTGCGCGAGCCCTGGCGGCGCTTACCACCGAC 10576
OY 427 GlnThrValIysLeuAlaIysThr-----ValIysAspMetSerGlyLeuAlaIleSerSer 445
DB 10575 AATGCAATCCGACATGAGAGACCAAGATCGGCGGATGTGTCGACACGCAATGATGCC 10516
OY 446 ThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProGlu--- 464
DB 10515 CTG-----ACCTCGCAAAACCTTGCCACCTTCTGTCGAGAGCGAGATC 10474
OY 465 -----AlaAspIleGluGluGly 471
DB 10473 AAGCAATCAACCAAGGCTCTGCGGCTGTGATGTCAACAAGTCAACACCGCGCAAC 10414
OY 472 IleSerAlaPheSerArgSerGlu 479
DB 10413 ATCGCGCGCTGACCAAGGCGAG 10390

RESULT 11
AE003482/C 301769 bp DNA 1linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386045 section 16
DEFINITION of 17, complete sequence.


```
/map="64B17-64C1"
/db_xref="FLYBASE:FBan0013715"
/db_xref="FLYBASE:FBgn0035564"
complement<11050..>11361"
/gene="CG13715"
/product="CT33181"
/evidence=not_experimental
/db_xref="FLYBASE:FBan0013715"
/db_xref="FLYBASE:FBgn0035564"
complement(11050..11361)
/gene="CG13715"
/note="CG13715 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AaF47927.1"
/db_xref="GI:7292525"
/db_xref="FLYBASE:FBan0013715"
/db_xref="FLYBASE:FBgn0035564"
/translation="MQKLPQRLSQGVGIQOOTSSKRRKPYGSGAGATVVGPA
TVRQVSMGPPIVAETDVLANNSSGNNGNKGKAKGKQNGHYVRRAITTR
N"
complement<19046..>49523)
/gene="axo"
/note="axo"
/map="64B17-64C1"
/db_xref="FLYBASE:FBan0018296"
/complement(join(19046..19786,20440..20558,23933..24392,
24553..24800,29589..29988,30176..30390,33096..33628,
33693..34025,37743..37939,38039..38561,43730..44019,
45440..45577,46789..47008,48572..48845,48998..49523))
/gene="axo"
/product="CTA1507"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/db_xref="FLYBASE:FBan0018296"
/db_xref="FLYBASE:FBgn0026729"
<21899..>23751
/gene="CG7474"
/map="64B17-64C1"
/db_xref="FLYBASE:FBan0007474"
/db_xref="FLYBASE:FBgn0035565"
join(21899..22134,22288..>23751)
/gene="CG7474"
/product="CT8014"
/db_xref="FLYBASE:FBan0007474"
/db_xref="FLYBASE:FBgn0035565"
22288..23751
/gene="CG7474"
/note="CG7474 gene product"
/codon_start=1

CDS
mRNA
gene

Alignment Scores:
Pred. No.: 1.56e+03 Length: 301769
Score: 146.50 Matches: 91
Percent Similarity: 39.32% Conservative: 61
Best Local Similarity: 23.70% Mismatches: 156
Query Match: 6.09% Indels: 77
DB: 3 Gaps: 13

US-09-825-414-66 (1-487) x AE003482 (1-301769)
QY 141 GGAAPTTPLeuAlaThrProleuysProleuThrProleuIleSerGlyAlaLeuSer 160
DB 271261 GCGGCGTTAGtGGCGAGGCTTCAGACGAGCGAGCGGTCTGTGAGCGGTAGCCATCGCT 271202
QY 161 GtAlaMetaspGlnValGlyThrIlysmetCaspArgAlaArgGlyAspLeuHis--- 179
DB 271201 GCGCAGAGACGTCGTCTGCGCGGACATCATGATGGAGAGCGGTCCTCCACGCG 271142
QY 180 -----TyrLeuSerThrSerProAspIlyLeu 188
DB 271141 GCGCAGTACTCGCCGAGAGCGCTCAGAGACATCACGCGCGGACGATCATCACAGACG 271082
```

```
QY 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGly 204
DB 271081 CACGCTCTTGTAGCGGTATTCATCATCACCCAGGGGAGTCTCCACAGACGCTCGGA 271022
QY 205 ArgGlnValAlaAspMetGlyTlleAlaValGln-----ThrPheSer 218
DB 271021 GCGGAGCTCTCTCAACAGGGGCGAGGTATTCAGGGGACAGCTCGAGACATCATCACAGACG 270962
QY 219 AlAlaAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238
DB 270961 ACGGATCACACGACGACGAGCTGTAGCGGTATTCATCATCATCAGCAGGAGATCTCGAC 270902
QY 239 GlyAlaValAspPheGlyValSerThrAlaGlyLeuValAlaAsnAlaIlePheGly 258
DB 270901 GGC-----AACGCTTCGGA 270887
QY 259 AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyValAlaPheValLeuGly 278
DB 270886 GCGGATTCCTC-----AACAGGGGCGAGGTATTCGGGGACAGCTCGTTAC--- 270839
QY 279 MetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeu----- 295
DB 270838 -----ATCACGACGCGCAGCGATCATCACGACGAGCATCTGTAGCGGTA 270794
QY 296 -----AspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
DB 270793 TCCATCAGCAGCAGAGGGGCTGTCAACACCAACGCTTCGAGGAGCACTCTTCAACGCGG 270734
QY 314 GlyLysArgMetAlaGlyLeuProLeu---AspValAlaThrAspGlyLeuLysAlaVal 332
DB 270733 GGGCAGCTCTCGCCGGAAGAGCTCATGTACATCACGACACGACGAGATCATCACGCGG 270674
QY 333 ArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu-----AlaLeuAla 349
DB 270673 GACGCTCTGTGATCAGGTACATCATCACAGAGGGGTCTCCACAGAACGCTTCAGA 270614
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
DB 270613 GCGGAGCTCTCTCAACAGGGGCGAGGTATTCAGGGGACAGCTTCGAAACATCACAGCAG 270554
QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGly---ValPhe 388
DB 270553 ACGGATCATACGCGCGGACAGTCTGTAGCGGTATTCATCATCATCAGGAGGAGTCTCGAC 270494
QY 389 AlaGlyTyrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPhe 408
DB 270493 ACCAACAGCCTCGAGGCGGACTCTCAACACAGGGGCGAGTACTCAGGGGAGAGATTCC--- 270437
QY 409 IleGlnAspLysValLysSerThrAlaSerSerThrLysSerTyrValAlaAspGlnThr 428
DB 270436 -----GGAGACATCACGACGACGACGAGATCACGCGGACGCTCTT 270395
QY 429 ValLysLeuAla-LysThrValLysAspMetSer-----GlyGluAlaIleSe 444
DB 270394 GTACCGGTATCATCATCAGCAGAGACTCTCGACAGCAGCGCTTCGAGAGCGGATTTC 270335
QY 444 rSerThr-GlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProG 464
DB 270334 CTCACAGGGGCGAGGTATT-----CACGAGGAGGAGCTC 270299
QY 464 LuAlaAspIleGluGluGlyIleSerAlaPheSerArgSerGluThrProPheGln 484
DB 270298 GTTCACATCATCAGCGGTGCGCACTTACAGCGGCGCAGCGTCTGTAGCGGTATTCATC 270239
QY 484 euArgArg 486
DB 270238 GTCCGAGA 270231
RESULT 12
LOCUS SME591788 306250 bp DNA linear BCT 05-JUL-2002
DEFINITION Sinorhizobium meliloti 1021 complete chromosome; segment 7/12.
```

ACCESSION	AL591788	AL591688	CDS	2887..4062
VERSION	AL591788.1	GI:15074579		/gene="SMC00955"
KEYWORDS				/function="miscellaneous: hypothetical/global homology"
SOURCE		Sinorhizobium meliloti.		/note="Product confidence : hypothetical
ORGANISM		Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobiaceae; Sinorhizobium.		Gene name confidence : hypothetical
REFERENCE		1 (bases 1 to 306250)		predicted by Codon_usage
AUTHORS		Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J., Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S., Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D., Pohl,T., Portetelle,D., Puhler,A., Purnelle,B., Ramsperger,U., Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.		predicted by Framed"
TITLE		Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021		/codon_start=1
JOURNAL		Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)		/evidence=not_experimental
MEDLINE		21396507		/transl_table=11
PUBMED		11481430		/product="CONSERVED HYPOTHETICAL PROTEIN"
REFERENCE		2 (bases 1 to 306250)		/protein_id="CAC46226.1"
AUTHORS		Gouzy,J.		/db_xref="GI:15074581"
TITLE		Direct Submission		/db_xref="SPRMBL:092PT3"
JOURNAL		Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELLIO EU Consortium		/translation="MLLYGACRPYGLVLEPAANPSAAGVEMLVTTTRSRREKGEFSGRALSPFAEMTVSIPDPYKAGVAPSKPLPPNETDFAFLKAVEVDRAEAKWLNTHVRSKSPGSGVLYVFIHGNNHFEDEAVFEFATIIHDSGRSPVLAATWPSRSLIAYGYDESTNTYRNAVERLFOYLARDEVEVSLIASHSGNMLATLSIOMATIRNDGLPAKFRNVMLEADVDVDFGSOIADMGKORPRTFLVSRDRLAVSRWGVNSRLAIDPCSPYKEELANGITVYDILTKIRAGNHLHTRFAESPEIVRLIGSRSLSSGOTLIDSRLGSHVIAATGAAHTVGTAGLVAAAPALIVDQTRNRYHHVEALAEPSGATR"
COMMENT		Laboatoire de Biologie Molculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes, France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetsstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequence.toulouse.inra.fr/meliloti.html.		4548..4724
FEATURES		Location/Qualifiers		/gene="SMC00959"
SOURCE		1..306250		4548..4724
		/organism="Sinorhizobium meliloti"		/gene="SMC00959"
		/strain="1021"		/function="miscellaneous: unknown"
		/db_xref="taxon:382"		/note="Product confidence : hypothetical
gene		488..2707		Gene name confidence : hypothetical
		/gene="SMC00954"		predicted by Codon_usage
CDS		488..2707		predicted by Framed"
		/gene="SMC00954"		/codon_start=1
		/function="SMC00954"		/evidence=not_experimental
		/note="Product confidence : hypothetical		/transl_table=11
		Gene name confidence : hypothetical		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		predicted by Codon_usage		/protein_id="CAC46228.1"
		predicted by Framed"		/db_xref="GI:15074583"
		/codon_start=1		/db_xref="SPRMBL:092PT1"
		/evidence=not_experimental		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/transl_table=11		5519..5737
		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"		/gene="SMC00958"
		/protein_id="CAC46228.1"		5519..5737
		/db_xref="GI:15074580"		/gene="SMC00958"
		/db_xref="SPRMBL:092PT4"		/function="SMC00958"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/note="Product confidence : hypothetical
		/codon_start=1		Gene name confidence : hypothetical
		/evidence=not_experimental		predicted by Codon_usage
		/transl_table=11		predicted by Framed"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/codon_start=1
		/protein_id="CAC46225.1"		/evidence=not_experimental
		/db_xref="GI:15074580"		/transl_table=11
		/db_xref="SPRMBL:092PT4"		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/protein_id="CAC46228.1"
		/codon_start=1		/db_xref="GI:15074583"
		/evidence=not_experimental		/db_xref="SPRMBL:092PT1"
		/transl_table=11		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"		5519..5737
		/protein_id="CAC46225.1"		/gene="SMC00958"
		/db_xref="GI:15074580"		5519..5737
		/db_xref="SPRMBL:092PT4"		/gene="SMC00958"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/function="SMC00958"
		/codon_start=1		/note="Product confidence : hypothetical
		/evidence=not_experimental		Gene name confidence : hypothetical
		/transl_table=11		predicted by Codon_usage
		/product="CONSERVED HYPOTHETICAL PROTEIN"		predicted by Framed"
		/protein_id="CAC46225.1"		/codon_start=1
		/db_xref="GI:15074580"		/evidence=not_experimental
		/db_xref="SPRMBL:092PT4"		/transl_table=11
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/codon_start=1		/protein_id="CAC46228.1"
		/evidence=not_experimental		/db_xref="GI:15074583"
		/transl_table=11		/db_xref="SPRMBL:092PT1"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/protein_id="CAC46225.1"		5519..5737
		/db_xref="GI:15074580"		/gene="SMC00958"
		/db_xref="SPRMBL:092PT4"		5519..5737
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/gene="SMC00958"
		/codon_start=1		/function="SMC00958"
		/evidence=not_experimental		/note="Product confidence : hypothetical
		/transl_table=11		Gene name confidence : hypothetical
		/product="CONSERVED HYPOTHETICAL PROTEIN"		predicted by Codon_usage
		/protein_id="CAC46225.1"		predicted by Framed"
		/db_xref="GI:15074580"		/codon_start=1
		/db_xref="SPRMBL:092PT4"		/evidence=not_experimental
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/transl_table=11
		/codon_start=1		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/evidence=not_experimental		/protein_id="CAC46228.1"
		/transl_table=11		/db_xref="GI:15074583"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/db_xref="SPRMBL:092PT1"
		/protein_id="CAC46225.1"		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/db_xref="GI:15074580"		5519..5737
		/db_xref="SPRMBL:092PT4"		/gene="SMC00958"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		5519..5737
		/codon_start=1		/gene="SMC00958"
		/evidence=not_experimental		/function="SMC00958"
		/transl_table=11		/note="Product confidence : hypothetical
		/product="CONSERVED HYPOTHETICAL PROTEIN"		Gene name confidence : hypothetical
		/protein_id="CAC46225.1"		predicted by Codon_usage
		/db_xref="GI:15074580"		predicted by Framed"
		/db_xref="SPRMBL:092PT4"		/codon_start=1
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/evidence=not_experimental
		/codon_start=1		/transl_table=11
		/evidence=not_experimental		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/transl_table=11		/protein_id="CAC46228.1"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/db_xref="GI:15074583"
		/protein_id="CAC46225.1"		/db_xref="SPRMBL:092PT1"
		/db_xref="GI:15074580"		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/db_xref="SPRMBL:092PT4"		5519..5737
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/gene="SMC00958"
		/codon_start=1		5519..5737
		/evidence=not_experimental		/gene="SMC00958"
		/transl_table=11		/function="SMC00958"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/note="Product confidence : hypothetical
		/protein_id="CAC46225.1"		Gene name confidence : hypothetical
		/db_xref="GI:15074580"		predicted by Codon_usage
		/db_xref="SPRMBL:092PT4"		predicted by Framed"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/codon_start=1
		/codon_start=1		/evidence=not_experimental
		/evidence=not_experimental		/transl_table=11
		/transl_table=11		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/protein_id="CAC46228.1"
		/protein_id="CAC46225.1"		/db_xref="GI:15074583"
		/db_xref="GI:15074580"		/db_xref="SPRMBL:092PT1"
		/db_xref="SPRMBL:092PT4"		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/codon_start=1		5519..5737
		/evidence=not_experimental		/gene="SMC00958"
		/transl_table=11		5519..5737
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/gene="SMC00958"
		/protein_id="CAC46225.1"		/function="SMC00958"
		/db_xref="GI:15074580"		/note="Product confidence : hypothetical
		/db_xref="SPRMBL:092PT4"		Gene name confidence : hypothetical
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		predicted by Codon_usage
		/codon_start=1		predicted by Framed"
		/evidence=not_experimental		/codon_start=1
		/transl_table=11		/evidence=not_experimental
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/transl_table=11
		/protein_id="CAC46225.1"		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/db_xref="GI:15074580"		/protein_id="CAC46228.1"
		/db_xref="SPRMBL:092PT4"		/db_xref="GI:15074583"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/db_xref="SPRMBL:092PT1"
		/codon_start=1		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/evidence=not_experimental		5519..5737
		/transl_table=11		/gene="SMC00958"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		5519..5737
		/protein_id="CAC46225.1"		/gene="SMC00958"
		/db_xref="GI:15074580"		/function="SMC00958"
		/db_xref="SPRMBL:092PT4"		/note="Product confidence : hypothetical
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		Gene name confidence : hypothetical
		/codon_start=1		predicted by Codon_usage
		/evidence=not_experimental		predicted by Framed"
		/transl_table=11		/codon_start=1
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/evidence=not_experimental
		/protein_id="CAC46225.1"		/transl_table=11
		/db_xref="GI:15074580"		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/db_xref="SPRMBL:092PT4"		/protein_id="CAC46228.1"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/db_xref="GI:15074583"
		/codon_start=1		/db_xref="SPRMBL:092PT1"
		/evidence=not_experimental		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/transl_table=11		5519..5737
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/gene="SMC00958"
		/protein_id="CAC46225.1"		5519..5737
		/db_xref="GI:15074580"		/gene="SMC00958"
		/db_xref="SPRMBL:092PT4"		/function="SMC00958"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/note="Product confidence : hypothetical
		/codon_start=1		Gene name confidence : hypothetical
		/evidence=not_experimental		predicted by Codon_usage
		/transl_table=11		predicted by Framed"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		/codon_start=1
		/protein_id="CAC46225.1"		/evidence=not_experimental
		/db_xref="GI:15074580"		/transl_table=11
		/db_xref="SPRMBL:092PT4"		/product="PROBABLE EXOEXOYRIBONUCLEASE III PROTEIN"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/protein_id="CAC46228.1"
		/codon_start=1		/db_xref="GI:15074583"
		/evidence=not_experimental		/db_xref="SPRMBL:092PT1"
		/transl_table=11		/translation="MKIATYNNVGNRGLVLLRWLEASPDVVCLOELKAPDKFPVYKALEAGYCAIHNGOKSNMGVALIARDEPRTLRKGLPGDDPDHSTRTIAAGMYITGCTLPNGNYPGKFEYKLAWEHRLTVAELLELDVPIVILAGDVNMPLELDVYFRERWWDALFRIEVDAYHRLDEGWDALROLHPGERVYTFMWYFRNAFARDAGLRIDHLLSPVTLRLSLAAGVDRVRGWEHTSDHAPAMIELSDGPAEDG"
		/product="CONSERVED HYPOTHETICAL PROTEIN"		5519..5737
		/protein_id="CAC46225.1"		/gene="SMC00958"
		/db_xref="GI:15074580"		5519..5737
		/db_xref="SPRMBL:092PT4"		/gene="SMC00958"
		/translation="MPAQPHIPANAPNIVIMLDVGPALPDTGCVTEPTLSRLAQDGLSYNFRHNAMCSPSRALLTGRNHRVNGOIAELANDWDYTGRIPTSTAVKVLGYGATAGAKMNTPANETTAGVPDPAGEIGEDYFEGFLAGESQSSQTSVAVVENTRVDPSSGKEGYHFTEDMTKATVMKQVAILPDRPFMYWAGAGAHNHFKDMAKYGKGGDDMDAMREIRYARQOLGMIPONTOLTPRPDTLSGMDIPEDEKFORRLMEVAGYTEHADVOAGRLIALDELIGRDTLIFYVWGNGSSASEONGITSEL		/function="SMC00958"
		/codon_start=1		/note="Product confidence : hypothetical
		/evidence=not_experimental		Gene name confidence : hypothetical
		/transl_table=11		predicted by Codon_usage
		/product="CONSERVED HYPOTHETICAL PROTEIN"		predicted by Framed"
		/protein_id="CAC46225.1"		/codon_start=1
		/db_xref="GI:15074580"		

```

/evidence-not_experimental
/transl_table=11
/protein_id="HYPOTHETICAL/UNKNOWN PROTEIN"
/db_xref="GI:15074584"
/db_xref="SPTREMBL:092Pr0"
/translacion="MTNMDRARLELRLEKLVDAAVAGARQDYAAIIIDEIGQLRL
AYEDPDPPADIDSEQYIEEESNMWPAAD"
5839..6005
/note="Sm-4 OR SMC04555
REPEAT SM-4
predicted by Homology"
/evidence-not_experimental
/complement(6137..9652)
/gene="mfid OR SMC04450"
/complement(6137..9652)
/gene="mfid OR SMC04450"
/function="macromolecule metabolism; macromolecule
synthesis, modification; dna - replication, repair,
rest./modif."
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE TRANSCRIPTION-REPAIR COUPLING FACTOR
(TREF) PROTEIN"
/protein_id="CAC46230.1"
/db_xref="GI:15074585"
/db_xref="SPTREMBL:092P89"
/translation="MTMIGCFDPKPTILATREITIGPVTGAELVLAELARAGAPYA
YIISDGKAVADLEQVLCGFVAPDIPVLTLPKGDCLPTVRPSADTSARRLAASALIA
HRRKFAIIVLVTINALQRIISPDVIESIAFTRAPENQIMDLARLERNGEVRAP
TRRMGEFAVARGGILVDVYVPGSEPRILDFEGDTELRISFEPASQRTICQVRLIDLN
PMSVSLPTETISHTFRKOYSLFGAARDDALVQAVSEGRYAGMEHMLPLEYRLDT
VFVDLDFGRIYTDHILAREAAERSEKILVDYDARLASASPGSQVGTGTPKVPYPM
LYIATAGFGEALDINAVRLSPTEHGEGAROVYNTAARGLRRARAGADNDGTT
NPDYAKHTEAKRRAKAKATVSGMTGSLDRLLQVLAEGHLANTIRYKALSDIGSLK
PGEASAVLSLEAGFETGDLVIVGEODILDDRILVRKRRKRGADFAEYTGDEGSY
VVAHEHIGRFGVGLRTIEAAGAPHDCLVYADAKLFLPVENTELSRVSGEGTDVAV
LDKRVGAQMARAKARKLKKRLDLMAAGLIRIAAERHTHAFLVLAODGVDEFARFPY
DETTELDQNSIDAVDDVDDGRGPRMDRLVCGVYGEFETVLAARAFLIAMNGGVAVVVP
DRLAROHKFTFSRFRGLPIRTOOASRLVGSMDLALTKREVAEGKDIYVGHALG
SSIKFAULGLIIDEBOHFVGVKKERKELKTDVHVUTLSNTPRLLQALGVREL
SLTTPKVPVDRMAVFTLSPEDALVIRTELRKREHNRGGQSYVCPRVSDLEPIHFLKS
DVPELKVAVAGQMPATELDEIMNAFVEGRYDVLTSTIVSGLDVPTANTLIVHRAD
MFGIAQLYQLRGVYGRSKVRAEALFTLPVKKTLTGPAERLRKVLOSIDTIGAGFOLAS
HDDIRGAGNLGLGEOSGRIKEVGFELYOOMLEAVAEIKGEEBIIHDTGSPQISVGT
PVMIPEYVPMINRLGLVYRLRGLGELTDLKEIDGALLIDRFGLPTEVYQHLKIVY

```

Db 184972	TCGGCGATGAACGAGGGCGATAGAGCCACCATCGCCCGGCGAACCGAGCTCGAAGGCGTC	185031
Qy 63	GlySerProasp.....	66
Db 185032	GTGCATTTCGGAAGTCACGGCGACTCGAAGCGAGCTACAGCGAAGAAACGAACGCGGTGTCGGC	185091
Qy 67	-----GluArgAspAlaIleuAlaHisasnGluInIle	78
Db 185092	ACGCGTCGCAGAGCTTGCTGCTCGAAGCGGAACCATCATCGCCATTCGGCATTCGGATC	185151
Qy 79	AspAlaIleuVal.....Glu	83
Db 185152	CGCAGCGCGATCGCTGGCGCGCATACCAACCTGAAGGAGCAGCCTCGAAGCGCAAGCG	185211
Qy 84	ThrArgAlaasnArgLeuTyrSerGluGluThrProAlaThrIleAlaGluThrPhe	103
Db 185212	GACATTGCGTGGCCGATCGCCGTCTCCGGCGAGCCCTTCGCTTCGCTCATCGATACCCGG	185271
Qy 104	AlaAlaAlaGluLysPheAspArgLeuAlaThrThrAlaSerAlaPheGluAsnThr	123
Db 185272	GCCGGCGCA-----CTTACCGACAAATCGGACCATCGCGCTCGAATAAT---	185313
Qy 124	ProPheAlaAlaIaIaSerValLeuGlnTyrMetClnProAlaIleAsnLysGlyAspTrp	143
Db 185313	-----	185313
Qy 144	LeuAlaThrProLeuLysProLeuThr---ProLeuIleSerGly-----	157
Db 185314	CTGAGCACCATGATGTAGCACCCGGACCGAGCGGCTTTCGGGCTTCGAGCCGCCGA	185373
Qy 158	-----AlaLeuSerGlyAlaMetAspGln	165
Db 185374	GTGGCGCTCAGACCAAGATTGAGCGGCGGTCTGCATCTCATCGATCAAGCATTCAGAGCAG	185433
Qy 166	ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrIleuSerThrSerPro	185
Db 185434	CGTGGCGAGCGAGTTTGCTCAGCGCAATTCGAACCCGTCCTCGACGCTCGAGCGCCATACC	185493
Qy 186	AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGluArg	205
Db 185494	GAGAAAGCTCAACGCCGCCCTGGAACGAAGCGCCCGCTGACGTCACACGAGACGCTGATCGCC	185553
Qy 206	GlnAlaValAlaMetGlyIleAlaValGlnThrPheSerAlaLeuAsnAlaValArgThr	225
Db 185554	CGAGACGCCGATTCACATGAGAGCGCTCAGG-----	185583
Qy 226	ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal	245
Db 185584	-----ATAGCGCACAGCGCATTTCTCCGGCGCCCTCGACAGCTGCTTT	185625
Qy 246	SerThrAlaGlyGlyLeuValAlaAlaAlaGlyPheGlyAspArgMetLeuSerValGln	265
Db 185626	TCCTCGCTCAACTCGGCCCTTGACGAAAGGAGCGAGCTTCGGCCAAAGCTTCAGACTCG	185685
Qy 266	SerArgAsp-----GlnLeuArgLysGlyAlaPhe-----	275
Db 185686	AGCGCGACAGCGGCGCATCATGATCTCGATCTCGCGGCGGCTTTTCGAGGAAAGCTG	185745
Qy 276	-----ValLeuGlyMetLysAspLysGluProLysAla	286
Db 185746	CAGACGAGCGTCGGCCAAATTTGGCAAGCGCCTTCGACGACGAGCTTCACGAAATTTGCACAGC	185805
Qy 287	AlaLeuSerGluGluThrAspTrpLeuAspAlaThrTyrAlaAlaIleLysSerAlaSerTyr	306
Db 185806	GCCTTTCGACAAAGCGGCGAGCGACGCTGACAC---AAGCTCAAGGAAAGCTTCACACGG	185862
Qy 307	SerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspAlaIaThr	326
Db 185863	ATCAACGAGACCGCTTTCGGGGGCTCGAGAGCCATCGCGGCGCCCTTGACACAGAGTGTG	185922
Qy 327	AspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerIleuThrLysAsnGlyLeu	346
Db 185923	GAC-----AAGATCAATTCGCGCCCTTTCGACGAGCTGCTCAC-----CTG	185964

FEATURES	source	gene
TITLE	Complete genome sequence of <i>Caulobacter crescentus</i> CB15	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)	
MEDLINE	21173698	
PUBMED	11259647	
REFERENCE	2 (bases 1 to 13040)	
AUTHORS	Nierman, W.C., Feldblyum, T.V., Paulsen, I.T., Nelson, K.E., Eisen, J., Heidelberg, J.F., Alley, M.R.K., Ohta, N., Maddock, J.R., Potocka, I., Nelson, M.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DebRoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M., Khouri, H., Shetty, J., Berry, K., Uterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolova, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L. and Fraser, C.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA	
FEATURES	Location/Qualifiers	
source	1..13040	
gene	/organism="Caulobacter crescentus CB15" /strain="CB15" /db_xref="taxon:190650" complement(260..2002) /gene="CC0921"	

complement(260. .2020)
/gene="CC0921"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="sensor histidine kinase/response regulator"
/protein_id="AAK22905.1"
/db_xref="GI:13422191"
/translation="NGMLAKAMSDERDEVAADTYRLTSLISALNAPALP
ALATWLSLTIGESLALLVTRKFRDGTGTPREKAMFAMASITPINISWTTLALLML
HEDMDLKLAAVAIWCGOVIYIQNFHQSAALLNLGLTPMASTLIPFEFEMPDASAA
QIARNGILVLYLVTTTNLWMIQNRRAARMDDELTRGIEREKALAEAAAKSFFIIVTSH
ELTRPMNGLIGAAHALDOSTLDPYQREDIALMISGSDLMOLDNDVLDLISVEFKV
LAPIMSLODIYEVVADMRAPAVKRNLSVDVAPLPCVHVAIDPLIRIQNVKYS
NAIKTVDCGVLEVAAPRANGDGRSITYADTGPYCPADAMKRVYDSFTQADQITIS
RHHGAGLGLSLARALAROMGGDVLVYSGAGCFVFEAPACAPAAAEVEYGANG
DALGEGYVIMADNAINOLDVYTMLEPGLAVTVENQOETLEMAERFCVLMIDP
MVMQGITAEIRDGTGNPAMPIVIALTASAMGDRERFGTGFDDHIGKPVKPVLD
ITAIKAPVPEPPNNGRRRA"
2258. .5278
/gene="CC0922"
2258. .5278
/gene="CC0922"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="medium-chain-fatty-acid--CoA ligase"
/protein_id="AAK22906.1"
/db_xref="GI:13422192"
/translation="MGLMONGALFLDKIIDHAAWHGGEVRSVSEGPVITTYGCO
IDBRKRVSNALNMGICSGDPVGLTANMGRHMDAWGINGICAVOCHTNRLPPEQO
IAMIADHADRITFDLTFPLIATLIRLEPHEVEYVFTTRAMPADRPACAGAPRPF
KCLICYELVEEQHADCCAMGGFDEGTAAGLCYTSGTGTDPKGVMSHNSNLHLTLTL
QDVGVLKSDCVLIPVPMFANAMGVAFSPGSCAKVMGAMKADASITYELDSQO
VYFSAAPVYQMLQHLAEATGATLPTLKATYIGACAPESIIIAFHDKYVEVVAAM
GMTESPVSGLTFLDELSEKLPEDQMMRKOGKRPVGVICGLDKDEBKLPHDGAAM
FENKTRIGRTITAEYFPGRAGKITDDEEFTPGDVAITDEXGFMQIDRAADKYSGG
ENITSTIEBNIAVGHPRKALAAVVGQPPHKKDERVYLVLKKEBETAPRPFIDYLOG
KIAKMMPPDVLIVDDIPLGATGKIDKLIRKLADDYKILPETHLOPSPAPPTLA
AADGSPKVIYAPPEVEEVAADVAHLTPEKAREETQDPBPVASEAISPEQPPRA
KILAEPEVOAKAPAPAPAEKAPVSAADAIITAEAAALIAAPKPAAEVTAEIKAPET
KASEVAKPMDAKSADADPLRLHEPDKAEPKAPRPVAAEPVAVDPGGRKARPL
KASDKKAKPKNASTYIDPAVYVATLPAIILAGCGALASVYKGLIDLPNGYNTMDMAR
KAAQGIATGLIGLVIALPAGFRLMKALALALTGTATFVWEGAKALRQOAPPHD
VATDKTPTLSEFSDAALAAAGTAAALVDDPSSLVPGSAAFGARLIAETNAETCPAARPL
VSSRSPADYAEKAAKAAVLASGMSLITDDPMQGRLEAFSGFMYGLITDYYVVRPDA
GSRIDRISGRDAGDGLGRNCRIGALMTAVKAR"
complement(5286. .6179)
/gene="CC0923"
complement(5286. .6179)
/gene="CC0923"
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="metallo-beta-lactamase family protein"
/protein_id="AAK22907.1"
/db_xref="GI:13422193"
/translation="MIPYRELFEFFYGRGDOVSLIRKVIARNGPFFETGTYIYG
REVAVIDQDLPREPHFKALASGEVYTVLIVHNHMDSLANPLAOLFKAUKYV
LPAEPDDEGAAPRSLEBDDARFRDYLADGEVLAGCGWLTETLANRSHNSLGN
LAEENALFEGDHIKGSSTVITTPRQDGGDIFASIAKRAAGFEFTLWPHCAQPREVVA
PRIDYIVVAHRRAREQOILEALGAGPTTIKAMPVSLYAADPRLPRAASHVLMIOQL
VREGSVVIAAGRDLAEYRLV"
complement(6513. .6920)
/gene="CC0924"
complement(6513. .6920)
/gene="CC0924"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK22908.1"
/db_xref="GI:13422194"

gene
CDS
/translation="MSDFREYTESLSVPOVDPADMAAAAGFVLVNNRPDGEDPG
QPSGSAEAAARAAGMDYLHVPRGGPDDVEIMRAFAEAAAGGPVLAFCSSGSRISV
TMSIGOLACGYDRETLVAOGYAGYDLSGVLP"
7526..10693
/gene="CC0925"
7526..10693
/gene="CC0925"
/note="similar to SP:P38370 GB:S64103 PID:402804;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="OmpA-related protein"
/protein_id="AAK22909.1"
/db_xref="GI:13422195"
/translation="MNSTKCGVRRRLSTLLAGLASIAAPLATTAVTAVPTLASAO
DYSSGTLGTTDNGAPVGAIVQVKSLSGFEKNVVGSDGFRILIPSGISVA
ISKAFVPASDGNVGVIGASNYGFIITAAAGVESELIVGKANPOLDFAGTGLS
IDVETLSKOVPIARSVNLTLAPSAVGDVGVFVNOGOQAANVGASIGENFEV
GLANTFENIGATLVPEEFYSVEIKGYPAGEGRTGAVINAVTSGSNDFTAL
HGNEPNSLRKSPNTITONSLSRDSRLIVEAGPILIKRLPEYGLASFOLETS
TATGGTYVDRTADPFYAFKLDGTYIDKHLELTPDTSRKRKRDPTMRADVIS
PTRASSELVYGGSENVYARTGTPTDMITVSAATGRSEVDAAVGNLSGARVDTR
TPTVRSIQRTAAGTSPSIAERETRADADLYELKLSHILRFIGDDENTLTPTTR
NGRNNYTRRAGAAGALGAVAAQOEYIEGRKFQYGGFAGKANAYITODAMDLRL
TSLGVRKDOQVADPNQOVSSFDNEIGLGCFTYDVGDRNDKLXGSGFRTYLPVA
SNTAFRAASPTADISEFMLPAGGCTTGALDAGTGIPAGIGRLVSGSLVACAPAY
SAAVGAFSCOVNRNNYSPNDVAARKLSTYDEYILGHEHPNDIMRGCVLIYR
NLGRISDALLDVCIQWNCANGAKYAKTNAATGCADVMSGHQYLLITPGSDVKNVLS
DPLGETOIRITIVSAQOLPFAKRETLGSEFSEFERTFDKMGLOGTYTISESKN
EGASRSDNGQTDAGITVDFDLAIIPGYGLLPHNRHQLLFPSTAVTDLLGLTNL
SVISPRKYGICGNAPENFDPNQDGLIEDQYEFANASYGAAAFCCGKAVNGSVFEGE
VTREDLSARYAVPSRFISLPYMDKSEGLVLADVFVNLKSGIEKDERGKTTGL
VNPFQKTYAVQRYVRFQDWSF"
complement(10690..11082)
/gene="CC0926"
complement(10690..11082)
/gene="CC0926"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK22910.1"
/db_xref="GI:13422196"
/translation="MGDGA RLKRRREADAVPSRLAFSGSLRMGALS LMPRPSP
LAGEVLRSRDEGLKALSGEARETPHPSPHPSAGRHLLPGKEEQJLASGISATVLSIRS
PSLRLFAKKRSGCIVAROSPRIR"
complement(11140..12924)
/gene="CC0927"
complement(11140..12924)
/gene="CC0927"
/note="identified by match to TIGR protein family HM
TIGR00705"
/codon_start=1
/transl_table=1
/product="signal peptide peptidase sppa"
/protein_id="AAK22911.1"
/db_xref="GI:13422197"
/translation="MKQFLLVAGVFCGLLVYGVPELLITAMAAGARAPVPAHAY
LQIDLRLNGLTQEPKSPPAAFSGSGSYTSITELRRAEKDKKALILVRLPEGVAP
AADELRLAEKHFREVGKPIYAHSQGLYPSGAVTSTMLAATSEFMQDSSFOAY
GISESEMFKEKFPYGVKVEYEQRYEKNVNPYLSHSDYPAHRESLTSMGSEYTR
ALTAAVDRKQDPTKMLKTELEGPYIAEOAOKGLIDVGQVSDIOAFLEKAGAK
LLDDDDYASRSKPTPAVSGPAIVAGIAGGATITGMDASAPSGDSNYSNDVQAPR
AATEDKVKATIVERVSSPGSDTASEOILALQAAKAGKRVVYVSKGYAASGQWIS
SQADSIITAPSTLUGSIGYIGKFRATIGALALRFVDRKDLHVGADYSQAFAGGFTP
EORAKFAGMDRITAGFTTRVAEGRNLPAPAVRIAGRWVTGQAKQLGVDELGGY
YEALEKAKALAKVDIRIKIMOGGASFEFEKFLVSETSAKTMAAAAVLDDPS
QAIDEMAKALRSTPGASVYLAADMPVY"
BASE COUNT 2203 a 4422 c 4259 g 2156 t
ORIGIN
Alignment Scores:
Pred. No.: 29.2 Length: 13040

Score: 145.00 Matches: 122
Percent Similarity: 36.14% Conservative: 58
Best Local Similarity: 24.50% Mismatches: 178
Query Match: 6.02% Indels: 141
Db: 1 Gaps: 24
US-09-825-414-66 (1-487) x AE005769 (1-13040)
QY 3 ILEASGLNSERIALGLINGLNPOROGILYALALAMETGLSERPHEARGTHALASER 22
Db 5305 CTGAATACCGCGCTCTGAGGACGCGCTCAGCGCGCGCTCATCAGGCGCGC 5246
QY 23 ASPALASERLEUALASERSESERVALARGSERVALSERTHTHRSERCYSARGASPLEU 42
Db 5245 GATCGCGGTGCGATGTCGGCGCCCAAGATCCGCGCTGCTCACCACGATGACCGGATATC 5186
QY 43 GINALAIETHPASPYRILEULYS-----HLSHSVALPHE----- 54
Db 5185 GAGCGCGCAACCGGCGCGCTCGGCGGCGACCCGCAACACACGATCGTGTGTCAGGCGGTA 5126
QY 55 -----ALALAHISARGPHE-SERVALILEGLY-- 63
Db 5125 CCAGAACTGTCGCGCTGCTCCAGCGGACCGGATCGATCATCTGATCAGGCT 5066
QY 64 -SERPROASGILUARGASPALAALALEUALHISANGLUGINILEASPALALEUVALGI 83
Db 5065 CATGCGCGACCGCAGGACGCGCGCTTGCGC-----GCCTCATAGGC 5024
QY 83 UTHRARGALAAANARGLEUTYRSERGLUYGLUTHRPROALATHRIEALAGLUTHRPH 103
Db 5023 GTGCGCGGCGAGCGGCTGTCACACGCGGTCGCGCGCGGACAGTCTCGCGCTGAT 4964
QY 103 EALA-----LYSALAEULUYSRPHASPARGLUEALATHRTHRAL 116
Db 4963 CTCGCGCATCGCGGACGCGCGGACGCGCGGACGCGGATGAGTGTGTCGTC 4904
QY 116 ASERSERALAPHEGLUASRTHRPHROPHALALA-----ALASE 129
Db 4903 GACCAAGCGC---GCCGTCGCCGCCCGCGCGGCGCGCGCTGGAGAACGAAAGCGG 4847
QY 129 RVALLEUNGINTYRMETGLNPROALAILLEASNYSELASPTRPLEUALATHRPROLEULY 149
Db 4846 CCTCTTCACAGTGGTGGCCACGTCGTGATCGCGGGCGCTGCGCAACAGCGCTTGGC 4787
QY 149 SPRO-----LEUTHPROLEULESERGLYALALEUSERGLYALAMETLS 164
Db 4786 CCCGAACATCACGCCAAGAGTCGCCCGGTGATCAGACAGCGCAAGAGCGCGCTTCCA 4727
QY 164 PGLN-----VALGILYTHRLYSMETHEASPARGALARGLYASPLE 178
Db 4726 -CAACCGGCAAGCGCGGCGGACAGCGCATACCAAGCGGATCAGCGCGCTGCGATCC 4668
QY 178 WHISTYR---LEUSERTHRSERPROASPIYLSLEULHISASPALAMETALA----- 193
Db 4667 CGGCTGGGCGCGCGCGGCGCGCGCATGTCATGATGTTGTAGCGCAAGCGCAGGT 4608
QY 194 -----VALSERVALIYSARGHISERPROALALEUNGILYAG----- 205
Db 4607 CGATCAGCGCCAGCTTGACGCCCAAGAGCGCGCGCGCGCGATCCGCCGGGTCAGGG 4548
QY 206 -----GLINVAL-VALASPMET-----GLYILEALAVAGINTHRP 217
Db 4547 CGACCAAGACCGCAAGTCCAGATATGATGACGCCAGCGGCGCTTGCGCTTCCGCGT 4488
QY 217 HESERALEULANSVALVALARGTHRVALLEUALPROALALEUALASERARGPROSERV 237
Db 4487 CCTCGCTTGGCGGCTC-----TTGCGCTTGGACCGCGGCGG 4452
QY 237 ALGINGLYALAVALSPPHEGLYVALISERTHRALAGLYLEUVALAASNALAGLYP 257
Db 4451 TGATCGGCAAGCGCGGCAAC---GGCGTGGCTCTCGCGC-----ACCAAGCGCGCGGCT 4404
QY 257 HEGLYASPARGMETLEUSERVALGINSERARGASGLNLEUARGGLYGLYALAPHEVAL 277

Db 4403 TCGGGGGCTCGGCTTGGGGCTTGGCAGC-----CGCAGCGCGCGC----- 4360
Qy 277 euglymellysAspLysgluProLysAlaAlaLeuSerGlulThrasPrlLeuAspa 297
Db 4359 -----TCGGCGCGC-----GACT 4347
Qy 297 laTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgm 317
Db 4346 TGGCGTCGGCTGGCCCTCGCGCGCTTCACCTCGAGACGCTGTCTCGGGGGCTTGA 4287
Qy 317 etalaglyLeuProLeuAspAlaIleThrasPglLysAlaValarSerLeuValS 337
Db 4286 TCTCGGGAGCGGTACCGCGCGCGGCTTGGCGCTCGCGCGCGATGGCGGCGCTT 4227
Qy 337 eraIa-----ThrasLeuThrasnglyLeuAlaLeuAlaGlyLysTyrAlaGlyV 355
Db 4226 CGGGGATGATCGGCTCGCGCGCGAGACCGGCTTGGCTCGCGCGCGCGCTCGGGGG 4167
Qy 355 alSerLysleuGlnLysMetAlaThryAsnIleThrasPserAlaThryLysAlaVal 375
Db 4166 CC-----TTGGCCTGACCTCAGCTCAGCGCGCGA 4137
Qy 375 alSerGlnLeuSerAsnLeuValGlySerVal----- 385
Db 4136 TCACCTTGGCTTGGCGC-----GGCGGCTTGGCGGCGAGATGGCTCGACGCGCGCA 4083
Qy 386 --GlyValPheAlaGlyTyrThrrAlaGlyLeuAlaThrasProlaValLysLysA 405
Db 4082 CGGGCTCGCGCGCGCGGCTTCTCTTGGCGCGCGCTCGCGCGCTGCGCAAGACAG 4023
Qy 405 laGluSerPheIleGlnAspLysValLysSerThrrAlaSerSerThrrThr 421
Db 4022 CG-----TCGGCGCGCAGCTCTTCGACG 4000

RESULT 14
AF067776 6413 bp DNA linear BCT 07-JAN-1999
LOCUS Abiotrophia defectiva extracellular matrix binding protein (emb)
DEFINITION gene, partial cds.
ACCESSION AF067776
VERSION AF067776
KEYWORDS AF067776.1 GI:3249002
SOURCE Abiotrophia defectiva.
ORGANISM Bacteria; Firmicutes; Lactobacillales; Aerococcaceae; Abiotrophia.
REFERENCE 1 (bases 1 to 6413)
AUTHORS Manganello, R. and van de Rijn, I.
TITLE Characterization of emb, a gene encoding the major adhesin of Streptococcus defectivus
JOURNAL Infect. Immun. 67 (1), 50-56 (1999)
MEDLINE 99081722
PUBMED 9864195
REFERENCE 2 (bases 1 to 6413)
AUTHORS Manganello, R. and van de Rijn, I.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Microbiology and Immunology, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA
FEATURES
source
1. 6413
/organism="Abiotrophia defectiva"
/strain="NVS-47"
/db_xref="taxon:46125"
248..>6413
/gene="emb"
/note="contains 17 tandem repeats of 231bp"
248..>6413
/gene="emb"
/note="adhesin"
/codon_start=1
/transl_table=11
/product="extracellular matrix binding protein"

/protein_id="AAD03320.1"
/db_xref="GI:3249003"
/translation="MLGKNFHLRLREGKRYTHAIKRLSGVASAVGSLFLGSS
MAQAOVVASQASTSAPASQGEQAOATVQASSSSOELESSESSQASLASEQA
LSSESSQSSSESSSSQSOSEPTPKPDPTROTNIPTPTMAYVGGGTALRDK
VERPATAMORPNNOGQPMPTGSLIGIGKGTDLIEPGAAIIVPDONLSNVK
RLKAAVIRLNPGRGITEGNIETTDGRLVITPKRSQGDYKSLYDVTDLAPPA
EPTVNGTIEDKLSSTDAITYGITPGATIKVTLQDMSVOTTRDRGNNTILKPKR
DSQYITGKGALIEVYQIMNGESAVERKTYDLEFTLLENLSSNRSRGKALDAQ
YIELAVBDALFMRSLATITROSITSMAYHGGGTALRDK
AKPLIKIEVLPGAQRODPTDSYHNVDFEIKKAGDYKKGKDRPGLTLPV
AALRIYRNVSGGVPEHWRMTSCNMAATPILVITPTSLYKNDALISDITISMYR
NDLEDDARQNTNGTPVQADIVYSKNGTYAIDPTYBETTYTLARSDSKSNMY
QVVAFTLPRERKAMVNNAKAKNTAIDNNNLTAIEKAEAKYAEAKNAATLACID
QAKTTAARNAAQKGTDTINAVPVPVAKPAAANALAEQAVNKINTEISQPDLTREK
QAMPDOYRTARDAAMAKVASANNQOATSRODGLAVNNLPPPAKYPALGIVRA
ADAKROAIRDNALTEEOADALROVDAQTAELAINONHTATLAKAPSDSVKIN
DINPQRSKRAANQALEOVAARQAI NNNOULTEDEKQAIQOYVOALANAKTYOA
ANDNGVNOAKTIGTATININIGQTOKAQALIAIEBAQKLEILOGRDLITERN
NALADITAKQAQAKDAVNOANNTGVGAARDQVADIGINPAVVKPDKRNIIDQA
RDEAEFOANTKLTDEKKAALIKVVOADARDAIDRAGSNDGVNNAVQGAALQA
IKALDQSPAKDTAKAIONADAKKAATTANNALTEDEKKAALIKOVDEKAAQAA
VDASRSKADYDRAKDOGLQISDVPVAVPEKLAIAVDOADTKRAVINOTTLQIE
EKRAALTRKYDEBAKARQAIINDATSNADVAKQAGCTQAINNTPOTPAKNAKAVE
QADAKKQAIENDPNLTRQEKDAIAKVDQETKARQAIIDATNDVTRAKQEGTQA
INAVPOTPAKTKAKNAVTOAADKSALENDNLTRQEKDAKVADEATKAKNI
DAATSNDEPRTAKQNEGTQAINAVPOTPAKTKAKNAVTOAADKSALENDNLTRQ
KVAKAKVADEAKKAKDAIDATSNADVTAKQNEGTQAINAVPOTPAKTKAKNAVTO
AADKKAIEKQNDLTRQEKDAKVADEATKAKDAIDATSNADVTAKQNEGTQAIN
NDVPOTPAKTKAKNAVTOAADKSALENDNLTRQEKDAKVADEAKKAKDAID
AATSNADVTAKQNEGTQAINAVPOTPAKTKAKNAVTOAADKSALENDNLTRQEK
DAKAKVADEATKAKNAIDATSNADVTAKQNEGTQAINAVPOTPAKTKAKNAVTO
ATDKKSAIENDPALTRQEKDAKVADEATKAKNAIDATSNADVTAKQNEGTQAIN
AVPOTPAKTKAKNAVTOAADKSALENDNLTRQEKDAKVADEAKKAKDAID
ATSNADVTAKQNEGTQAINAVPOTPAKTKAKNAVTOAADKSALENDNLTRQEK
AVKAKVADEAKKAKDAIDATSNADVTAKQNEGTQAINAVPOTPAKTKAKNAVTO
TDKKNATENDPALTRQEKDAKVADEATKAKDAIDATSNADVTAKQNEGTQAIN
VPOTPAKTKAKNAVTOAADKSALENDNLTRQEKDA"

BASE COUNT 2136 a 1592 c 1087 t
ORIGIN

Alignment Scores:
Pred. NO: 14.1 Length: 6413
Score: 143.50 Matches: 123
Percent Similarity: 36.22% Conservative: 78
Best Local Similarity: 22.16% Mismatches: 249
Query Match: 5.96% Indels: 105
DB: 1 Gaps: 23

US-09-825-414-66 (1-487) x AF067776 (1-6413)

Qy 3 IleaSnglInserAlaGlnPro-----ProGlyValAlaMetGlu 16
Db 2423 ATTAATGAATACAGCAACGCGCAGATTGACGTGTGAAGAGCAAGAGCTTCATGAT 2482
Qy 17 SerPheargThrrAlaSerAlaSerLeuAla-----SerSerSerValArgSerVal 34
Db 2483 CAGGTACGAGCTGCACGCGGCGGTATGGCAGAGTTGGCACGGCGGCGCAACATCAG 2542
Qy 35 SerThrThrSerCysArgasp-----LeuGlnAlaIle-----ThrasPtyr 48
Db 2543 GCCGTGACCAATGCCCGCGACCAAGGCTGAATCGGCAATATTACCAACACCGACA 2602
Qy 49 LeuYshIshIshValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArg 68
Db 2603 GCCAAGTACCCAGAGCGCTTGGCCAT-----GAGCCCAAGCAGCGCATGCTAAG 2653
Qy 69 AspAlaAlaLeuAlaHisAsn-----GluGlnIleAspAlaLeu--- 81
Db 2654 CGTCAGGCTATTCCGACCAATGCCACCTTACACCGCAAGAACACAGCGGCTCGCT 2713
Qy 82 -----ValGluThrArgAlaAsnArgLeuTyrSerGluGluThrProAla 97

```

Db 2714 CAGGTAGATGTCGCCAAACAGCGGCTGAAGTGCATTAAACCAAGAACATACCAAGCA 2773
QY 98 ThrIleAlaGluThrPheAlaIylAspGluLysPhe-----Asp 110
Db 2774 ACCTTGCTAAAGGCTGACATGATGGGCTTAAGCTAATTAATGACATTAAATCCGCAACCA 2833
QY 111 ArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaIleAlaSerVal 130
Db 2834 AGAAGCAACCAAGCTGCGCAACCAAGCTTTAGACCAAGTA-----GCAGCAGCCAG--- 2884
QY 131 LeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeu-----Ala 145
Db 2885 -----CGTCAAGGCATCAACAAATTAATGACATTGACAGGTGAAGAAGAGCGC 2932
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyVal 162
Db 2933 CAAGCCATTCAGCAGGAGGACCAAGCTTTGGCTTAATGCCAAGCTCAAGTTCAAGCTCT 2992
QY 163 MetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSer 182
Db 2993 AAGGACAAATTAAGCGGCTTAACCAAGCTTAAGACGCGC---GGGACAAACCGCTATTAAACAC 3049
QY 183 ThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAla 202
Db 3050 ATTAATCTCTCAAGGACCCAGCAAGCGGCAAGCCATTGCAATTTGAAGCGGCTGAACAA 3109
QY 203 LeuGlyArgGlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnVal 222
Db 3110 GCCAAACGCTTGAATTTGAGGGTCCGMAATGATTTGCCACCAAGCCCAATTAATGCC 3169
QY 223 ValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln----- 238
Db 3170 CTGGCAGACTTAAGCGCTAAGCGGCAAGCGGCCAAAGACCGGCTCAACCAAGCAGCAAC 3229
QY 239 -----GlyAlaValAspPheGlyValSerThrAlaGlyLysLeuValAla 253
Db 3230 AATACGGAGATGTCGGGTGCCAAAGATTAATGTGTGCTCAAAATTCAGAGGATTAATCT 3289
QY 254 AsnAla-----GlyPheGlyAspArgMet 261
Db 3290 ACAGCTGCTCTAAACCAAGCAGCTCGAATGCTATCGACCAAGCGGCTCGTGAATAAGAA 3349
QY 262 LeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysASP 281
Db 3350 GCAGATTCAGCGCTAATTAAGTTC-----ACAGAT 3382
QY 282 LysGluProLysAlaAlaLeuSerGlnGluThrAspTrpLeu---AspAlaTyrLysAla 300
Db 3383 GAGAGAAAGCAGCGCGCATCAGAAAGTTCAAGATCCCGCGGTATGCCAAGCGCGCC 3442
QY 301 IleLysSerLysArgTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMet 317
Db 3443 ATTGATCGACTGTTCTTAATGCGCATGTCACAAATGCCCTCAACCAAGCTAAG----- 3496
QY 318 AlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal---ArgSerLeuVal 336
Db 3497 -----GCAGCTATCCAAAGCCATCAAGCCCTAGATGACAGCGACCA 3538
QY 337 SerAlaThrSerLeuThrLysAsnGlyLeu-----Ala 347
Db 3539 AGTGCCAAAGATACAGCAAGCAGCATCCAAATGCCGAGATGCTAGAAAGCAGCC 3598
QY 348 LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThr 367
Db 3599 ATTACAGCCAAACATGCTTTGACTCAAGAAAGAAAGACAGCGCATTAAGCAAGTGAA 3658
QY 368 AspSerAlaThrLysAla---AlaValSerGlnLeuSerLysLeuValGlySerValGly 386
Db 3659 GATGAAGCAGCAGGCTCAAGCAGCGTAGACGCTTCAAGATCCAAAGCCGAGCTAGAC 3718
QY 387 ValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
Db 3719 CGGGCCAAAGACCAAGGCTCCAAAAAATTAAGCATGTGCGACGCTGCAACCAAG 3778

```

```

QY 407 SerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAsp 426
Db 3779 CTCATATGCCATTCGCGCAGATGATCAAGCGGCGATCAAGAGAGCCGTGATCAACAT 3838
QY 427 GlnThrValLysLeuAlaThrValLysAspMetSerGlyLysAlaIleSerSerThr 446
Db 3839 GATACGAGCT---TTGACCCAGGAAGAGCAAGAGCTCTATCCGCAAGTTATGAAGAA 3895
QY 447 GlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAsp 466
Db 3896 GCAGCCAGGCGCGTCAAGCGCATCAATGATGACAGCAGCAATGCGGATGTAGCGGTAAG 3955
QY 467 IleGlnGlnGlyLysLeuSerAlaPheSerArgSerGluThrPro 481
Db 3956 CAAGCCCAAGGACCAAGCCATTAAACAAATGTGCCAAACCAACCA 4000

RESULT 15
AC016584/c 132457 bp DNA linear PRI 29-MAY-2002
LOCUS AC016584
DEFINITION Homo sapiens chromosome 19 clone CTD-2596015, complete sequence.
ACCESSION AC016584
VERSION AC016584.5 GI:21240686
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 132457)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE 2 (bases 1 to 132457)
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Direct Submission
SUBMITTED (04-DEC-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 132457)
DOE Joint Genome Institute and Stanford Human Genome Center.
REFERENCE Direct Submission
AUTHORS Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell
JOURNAL Drive, Walnut Creek, CA 94598, USA
On May 29, 2002 this sequence version replaced g1:12965319.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.7% of Sequence;
Estimated Total Number of Errors is 0.3.
NOTE: This insert is not the entire sequence of the clone (entire
sequence is 138.1kb). It is clipped at the overlap with AC011464.
The number of bases overlapped is 14325.
Location/Qualifiers
1. 132457
source /organism="Homo sapiens"
/ab_xref="taxon:9606"
/chromosome="19"
/clone="CTD-2596015"

BASE COUNT 34214 a 30343 c 33402 g 34498 t
ORIGIN

Alignment Scores:
Pred. No.: 859 Length: 132457
Score: 143.00 Matches: 113
Percent Similarity: 33.67% Conservative: 87
Best Local Similarity: 19.02% Mismatches: 190
Query Match: 5.94% Indels: 204
DB: 9 Gaps: 25

US-09-825-414-66 (1-487) x AC016584 (1-132457)
QY 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSer 22
:::||||| |||:::

```

```

Db 12307 GTGAACACGAGACTTCAGAG-----AGAACACATCCAGGTGCCACG 12263
QY 23 AspAlaSerLeuAlaSerSerSerValArgSerValSer-----ThrThrSer 38
Db 12262 TCTGTACAGACAGACGCTCTCATTCACCATGTGTCTAGTGTGATCTCTCAT 12203
QY 39 CysArgAspLeuGlnAlaIle-----ThrAspTyrLeuLysHisValPhe 54
Db 12202 GTACACCAAGACAAAGCCACTTCTCTACGGAGACATCCATGCAAGCCCTCATCGTTT 12143
QY 55 -----AlaAlaHisArgPheSerValIleGlySerProAspGlnArgAspAlaIle 72
Db 12142 ATAACCTTCACCAACACATTCAGATGTAGACACCAACCCCTCCACTCTGTGATAAG 12083
QY 73 AlaHisAsnGlnGlnIleAspAlaLeuValGlnThrArgAlaAsnArgLeuTyrSerGlu 92
Db 12082 ACAGAAATCTTCAGAGGTGACCATCCACCACCAAGGTCTCTACTGTGAGCTGCACACG 12023
QY 93 Gly-----GlnThrProAlaThrIle 99
Db 12022 GGTCCATATCTCTGTGACATCAACCAACCATGCTTACTTGACAGAGCTCATTTAGCTGTG 11963
QY 100 AlaGlnThrPheAlaIysAlaGlnLysPheAspArgLeuAla----- 113
Db 11962 ACTCCAGATTTTATGCAATCAAGAGAACCTCTCATAGCAAAAGTCCCAAGATGTG 11903
QY 114 -----ThrThrAlaSerSerAlaPheGlnAsn-----ThrProPhe 125
Db 11902 TCCTGTGACAGAGCCCTCCCTGTGCGAGAAACAGCATACCTCTTCCCTGCACCTTTC 11843
QY 126 AlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsn-----LysGlnAspTyr 143
Db 11842 TTGGTCAACACCAT-----CCCTCGCCACATTCACAGCTTACAAAGGCCAACAT 11795
QY 144 LeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMet 163
Db 11794 ACATCTCTCTCGTTTCTCGCACTTCAGTTCTTACCTGTGGACTGGTG----- 11747
QY 164 AspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThr 183
Db 11746 -----AAGACCAAGATATGTGAACACAAAGCATGGAACCTGTGACCAAT 11702
QY 184 SerProAspLysLeuHisAsp-----AlaMetAlaVal 194
Db 11701 TCACCTCAAAATTTGAAACATTCATCAATGAGATAGTGGCCACTTGGCAGCCACACA 11642
QY 195 SerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaVal 214
Db 11641 GATATAGAGACTATTCATCTTCCATTAACAAAGCAGCATGACCAATATGAGGG----- 11591
QY 215 GlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArg 234
Db 11590 ---ACTGCCAGTTGACACATGACTGATGCATTCACACTCTC-----CCAGTCAAGCTCAGAA 11540
QY 235 ProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsn 254
Db 11539 CCATCT-----ACAGCCACATCTCCATGGTTCTCTGCC 11507
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAla 274
Db 11506 TCCAGCATGGGGGACGCTTGTCTTATATCA----- 11474
QY 275 PheValIleuGlyMetLysAspLysGlnProLysAlaAlaLeuSerGlnGlnThrAspTyr 294
Db 11473 ---ATACCTGGT-----TCTGAGACCAAGAGC--- 11450
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 11450 ----- 11450
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaIleThrAspGlyLeuLysAlaValArgSer 334
Db 11449 -----ATTGAGGAGAGCCAACTCTCTGCTGACTGTGAGAAAGAAAGAACAGCACC 11396

```

```

QY 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGly 354
Db 11395 CTCCAGGAGATGAACATCAACTACAGATGCAAGTCAACATCATCTCTCCAAATGTGTGTGGG 11336
QY 355 ---ValSerLysLeuGlnLysMet----- 361
Db 11335 GCTATTACTGAGCCACAAAAATGGAAGTCCCTCTTTGATGCAACATTCATACCAACT 11276
QY 362 -----AlaThrLys----- 364
Db 11275 CCTGTCACTCAAAAGATTCCTCCAGATATTTTCTCAGTAGCCAGAGTAGACTTTCAAC 11216
QY 365 -----AsnIleThrAspSerAlaThrLysAlaAla 374
Db 11215 TCTCTCCCATGACAATATCTACCCACATGACACCCACCCAGACAGGCTTCTGTGAGCT 11156
QY 375 ValSerGlnLeu-----SerAsnLeuValGlySerValGly----- 386
Db 11155 ACATCAAGATTCACCTTCCTTGAACATCAACCTTGGAAACCTCAGACAGGACTCCA 11096
QY 387 -----ValPheAlaGlyTyr-----ThrThrAlaGlyLeuAlaThrAsp 399
Db 11095 TCAGTGTGACTGAGAGGCTTGCACACTCAAAATPACCACCTGCAATGAAACATATGTC 11036
QY 400 ProAlaValLysLysAlaGlnSerPheIleGlnAspLysValLysSerThrAlaSer--- 418
Db 11035 AAGAGAGTGTACACAGCAAAACCTCCCTTTCAGATGAAAGCAGCTCCCTCTCTCA 10976
QY 419 -----Ser 419
Db 10975 GCACCTGCTCTGTGCAACAACCTTACCTTCTGTGCTTTCACACGCCAATGGCAGAGT 10916
QY 420 ThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLys----- 436
Db 10915 ACCTCTCTCTCGTTTCTATGTGCTTCAGTCTTCTTACTTCTTCACTGTAAGGCCAGGC 10856
QY 437 -----AspMetSerGlyGlnAlaIleSerSerThrGlyAlaSerLeuArgSerThrVal 454
Db 10855 AAGGTGATACAAAGCTTGAAGAACAGTAGACAGTCACTCAAAAGTATGAGCAACTTGG 10796
QY 455 AsnAsnLeuArgHisArgSerAlaProGlnAlaAspIleGlu 468
Db 10795 GATGACATATCGTGCACTTCACGACCCACACAGATATAGAG 10754

```

Search completed: January 31, 2003, 06:35:09
 Job time : 3603.21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:18:04 ; Search time 239.246 Seconds

(without alignments)
4584.089 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407

Sequence: 1 MHINSAQOPPCVAMESFRT.....EEGISAISRSRFPQLRRL 487

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame-p2n.model -DEV=rlp
-Q=cgn2.1/USPTO.spool/US09823414/rnatc.28012003.161741.26075/ppp_query.fasta.1.1294
-DB=N.Geneseq.101002 -OPMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsu62 -TRANS=humana4.0.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09823414.ecgn.1.1.103.etrnatc.28012003.161741.26075 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-MARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2407	100.0	1464	22	AAD20438	Pseudomonas syringae
2	1869.5	77.7	1461	22	AAD20408	P. syringae pv. tom
3	1869.5	77.7	1461	22	AAF55683	Nucleotide sequenc
4	1869.5	77.7	30365	22	AAD20405	P. syringae pv. to
5	148.5	6.2	5397	23	ABL29757	Drosophila melanog
6	147	6.1	4645	24	AAD31881	Drosophila melanog
7	146.5	6.1	3146	23	ABL23378	Drosophila melanog
8	146.5	6.1	9542	22	AAD04029	Moraxella catarrha
9	142	5.9	7029	23	ABL29756	Drosophila melanog
10	140	5.8	7100	20	AAK24980	Bacillus subtilis
11	138	5.7	3690	23	ABL29839	Drosophila melanog
12	138	5.7	3831	22	AAF67769	Corynebacterium gl
13	138	5.7	6136	22	AAF67767	Corynebacterium gl
14	138	5.7	6169	23	ABL29838	Drosophila melanog
15	138	5.7	8979	22	AAH67704	C glutamicum codin
16	138	5.7	349960	22	AAH68532	C glutamicum codin
17	136.5	5.7	4402	23	AAH89717	DNA encoding novel
18	136	5.7	3300	15	AAO57972	rsaa gene. Caulob
19	135	5.6	2708	23	ABL19629	Drosophila melanog
20	135	5.6	18234	22	AAH62719	Shrimp white spot
21	135	5.6	305107	22	AAH62689	Shrimp white spot
22	132.5	5.5	1230025	20	AAK91990	Nucleotide sequenc
23	132	5.5	3300	17	AAV17717	Caulobacter rsaa g
24	132	5.5	3300	18	AAV01866	Caulobacter cresce
25	132	5.5	3300	21	AAZ50079	Caulobacter cresce
26	131	5.4	1230	18	AAH67029	Salmonella secreta
27	130.5	5.4	7104	23	AAH51998	Staphylococcus aur
28	130.5	5.4	7107	23	AAH54654	Staphylococcus aur
29	130	5.4	12036	11	AAQ04668	FHA structural gen
30	129.5	5.4	6228	23	AAH55178	Staphylococcus aur
31	129.5	5.4	6561	23	AAH55178	Staphylococcus aur
32	129	5.4	7035	23	AAH54878	Staphylococcus aur
33	129	5.4	7240	22	AAH504273	Rat glutamate tran
34	128.5	5.3	3606	23	ABL09625	Drosophila melanog
35	128.5	5.3	9345	23	ABL09624	Drosophila melanog
36	127.5	5.3	7434	23	AAH552179	Staphylococcus aur
37	127.5	5.3	7437	23	AAH55232	Staphylococcus aur
38	127.5	5.3	4403765	22	AAH99683	Mycobacterium ture
39	127	5.3	349960	22	AAH68525	C glutamicum codin
40	126.5	5.3	2035	23	ABL10547	Drosophila melanog
41	126.5	5.3	32768	19	AAV52204	Streptococcus pneu
42	126.5	5.3	4411529	22	AAH99682	Mycobacterium tube
43	125.5	5.2	1669	21	AAH08793	Fusarium venenatum
44	125.5	5.2	4197	16	AAO99430	B. sphaericus SLP
45	125.5	5.2	48531	24	AAH20800	Clostridium difficile

ALIGNMENTS

RESULT 1	
AAD20438	
ID	AAD20438 standard; DNA: 1464 BP.
AC	AAD20438:
XX	
DT	03-JUN-2002 (first entry)
DE	Pseudomonas syringae pv. tomato strain DC3000 HopProA2 DNA.
XX	
KM	Conserved Effector Loc1: CEL; cytosolic; antibacterial; gene therapy;
KW	Exchangeable Effector Loc1: EEL; disease resistance; transgenic plant;
KM	eukaryotic cell death; cancer; ds.
XX	
OS	Pseudomonas syringae.
XX	
FH	Key
FT	CDS Location/Qualifiers
	1..1464

|||||
DB 1201 GCGGTAAAGAACGCCAGCGTTTATACAGATTAAGCTAAATTCACCCCATCTACTAC 1260
QY 421 ThrSerTrValAlaAspGlnThrValIysLeuAlaIysThValIysAspMetSerGly 440
DB 1261 ACMACTATGTGTCCGACACGACCGCTCAAACTGCGAAACAGTCAAGACATGAGCGCG 1320
QY 441 GluAlaIleSerSerThGlyAlaSerLeuArgSerThValAlaAsnLeuArgHisArg 460
DB 1321 GAGGCGATCTCCAGACCGCGTCCAGCTTACGCTACTGTCATTAACCTGCGTATCGC 1380
QY 461 SerAlaProGluAlaAspIleGluGlyIleSerAlaPheSerArgSerGluThr 480
DB 1381 TCCGCTCCGAGACTGATATCCAAAGAGTGGGATTTCCGCTTTCTCGAAGTGAACA 1440
QY 481 PropheGlnLeuArgArgLeu 487
DB 1441 CCGTTTCAGCTCAGCGCGTTTG 1461
RESULT 2
AAD20408
ID AAD20408 standard; DNA; 1461 BP.
XX
AC AAD20408;
XX
DT 03-JAN-2002 (first entry)
XX
DE P. syringae pv. tomato (Pto) DC3000 CEL open reading frame 5 (ORF5) DNA.
XX
KW Conserved Effector Loci; CEL; cytosolic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer; ds.
XX
OS Pseudomonas syringae.
XX
FH Key Location/Qualifiers
FT CDS 1..1461
ET /product= "Pto DC3000 CEL ORF5 protein"
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PE 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR.) CORNELL RES FOUND INC.
PA (UNRE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UNRE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR MPI: 2001-639361/73.
XX
P-PSDB: AAE12573.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
XX genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 1: Page 21, 217pp: English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant

CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 DNA.
XX
SQ Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other:

Alignment Scores:
Pred. No.: 9,59e-148 Length: 1461
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
DB: 22 Gaps: 1
US-09-825-414-66 (1-487) x AAD20408 (1-1461)
QY 1 MethisIleAsnGlnSerAlaGlnGlnPropGlyValAlaMetGluSerPheArgThr 20
DB 1 ATGCACATCAACCGACCGCTCCACACACCGCTTGACTGACCGATGACTTTGGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThSerGlyArg 40
DB 61 GCGTCCGACGCGCTCTGCTCCAGCTGTCGATGTCGATGTCGATGTCGATGTCGATGTCG 120
QY 41 AspleuGlnAlaIleThrAspTrpLeuIshIshValPheAlaAlaHisArgPheSer 60
DB 121 GAGATTAATGCGATTGCCGATTAACCTGACAGATGTTGGTGGCGATTAATGGCG 180
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
DB 181 CCGGCGGATTCGGCTGATGCGCAAGCTGCACTGACACATCAACATGCGCACTGCG 240
QY 81 LeuValGluThrArgAlaAsnArgLeuTrpSerGluGlyGluThrProAlaThrIleAla 100
DB 241 CTGATCGAGACGCGCGCGCGCTGCACTGCAAGGAGGAAACCGCGCAACCATCGCC 300
QY 101 GluThrPheAlaIleAlaGluIshPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
DB 301 GACACCTTCCGACGAGGAGGAAAGCTCGACGATTTGGCGACGATCAACGAGCGCTTG 360
QY 121 GluAsnThrProPheAlaAlaAlaSerValIleGlnTrpMetGlnProAlaIleAsnIys 140
DB 361 CCGGCGGACCGCTTGGCTCCATGCGCTGCTTCACTGATCACTGACCTGCTGATCAACAAG 420
QY 141 GlyAspTrpLeuAlaThrProLeuIshProLeuThrProLeuIleSerGlyAlaLeuSer 160
DB 421 GCGGATTTGGCTCCGCGCTCCGCTCAACCGCGACCGCTCATTTCCGAGCGCTGTCG 480
QY 161 GlyAlaMetAspGlnValGlyThrIshMetMetAspArgAlaArgGlyAspLeuIshIsh 180
DB 481 GCGGCAATGACCAAGGTGGGCAACAGATGATGAGCCGCGGCGATCTGCAATTGAC 540
QY 181 LeuSerThrSerProAspIshIshAspAlaMetAlaValSerValIshArgHisSer 200
DB 541 CTGACGCGCTCCGCGACAGCTCACATGATGCGCGCTTGGTGAAGCCACATCG 600
QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
DB 601 CCAGCCTTGTCTGACAGGTTCTGACACGCGGCGTTGCGGTTCAAGCATCTGCGCGCG 660


```

QY 101 GIuThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 301 GACACCTTCGCCAAGCGGAAAAAGCTCGACCGATTGGCAGACGATACACAGCCGCGTTG 360
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 CGGGGAGCGCCCTTTGCCATGGCTCGTTGCTTCACTACATCAGCCTCGATCAACAAG 420
QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 421 GCGGATTGGCTCGCGGCTCCGCTCAACCGCTGACCCCGCTCATTTCCGAGCCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
Db 481 GCGGCGCATGACCAAGTGGGACCAAGATGATGACCGCGGCGATGATCGATTGCAATTAC 540
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 541 CTGAGCGCCTCGCGGACAGGCTCCACGATGCGATGCCGCTTCGGTGAAGCCCACTCG 600
QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 601 CCAGCGCTTGCTCGACAGGTTCTGGACACGCGGGTGGCGTTGCAGACGTACGCGCGCC 660
QY 221 AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
Db 661 AAGCGCGGTACGATCGGATTTGGCTCGCGGATCGGCTCCAGACCGCCCTGAGGGTGT 720
QY 241 ValAspPheGlyValSerThrValGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
Db 721 GTGAGCACTTGTTGATTCATGATGGCGGTGGTGGCTGCCAAGCAGCGCTTGGCAACGC 780
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
Db 781 CTGCTCAGTGTGCACTCGCGGTATCACCAGCGTGGCGGGTGGCATTTGCTCGGTTGAAG 840
QY 281 AspLysGluProLysAlaAlaLeuSerGlnGluThrAspTrpLeuAspAlaTyrLysAla 300
Db 841 GATTAAGAGCCCAAGGCTCACTGAGCGAAGAAAGAGCTGCTGAGGCTTATTAAGCA 900
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 901 ATCAATATCGGCGCAGCTACGCGGTGCGGCGCTCAACGCGTGGCAGGATGGCCGCTCG 960
QY 321 ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
Db 961 CCACGTGATATGCGACGACGCAATGGGTGGGTAAAGAGCTGTGTCCACGCTCCACG 1020
QY 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyLysTyrAlaGlyValSerLysLeuGlnLys 360
Db 1021 CTGACCCAAAGCGTCTGGCCCTGGCGGGTGGCTTTCAGGGGGTGAAGAGTTGGCGAG 1080
QY 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 1081 ATGGCGACGAAAAATATCACCGACCGCGGACCAAGCGCGGTGAGTCAAGTCAACAC 1140
QY 381 LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
Db 1141 CTGGCAGGTTCCGACCGCTTTTCAGAGCTGACACCGCGCGCTGTCAACCCGATCCC 1200
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
Db 1201 GCGGTGAAAAAGCCGAGTCTTCATACAGCACGAGGTGAATTCGATCTCAATCCATACC 1260
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 1261 AGAGGCTACGTAGCGACAGACCGTCAAACTGGGGAAGACGCTCAAAACATATGGCGGG 1320
QY 441 GluAlaGlnSerSerThrGlyLysSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 1321 GAGGGATACCACTACATACCGCGCCAGCTTGGCCCAATACGCTCAATAAACCTGCTCAACGC 1380

```

```

QY 461 SerAlaProGluAlaAspIleGlnGlyIleSerAlaPheSerArgSerGluThr 480
Db 1381 CCGGCTCGTGAAGCTGATATAGAGAGGGGCGCGCGCT--TCTCAAGTGAATA 1437
QY 481 PropheGlnLeuArgArg 486
Db 1438 CCGTTTGGCCATGCGG 1455

RESULT 4
AAD20405/c
ID AAD20405 standard; DNA; 30365 BP.
XX
AC AAD20405;
XX
DT 03-JAN-2002 (first entry)
XX
DE P. syringae pv. tomato (Pto) DC3000 Conserved Effector Loc1 (CEL) DNA.
XX
KW Conserved Effector Loc1; CEL; cytosolic; antibacterial; gene therapy;
KW Exchangeable Effector Loc1; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer; ds.
XX
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR ) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
XX
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO:
XX
DR WPI; 2001-639361/73.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector loc1 and Exchangeable Effector Loc1
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Disclosure; Page 9-17; 217pp; English.
XX

The invention relates to an isolated nucleic acid molecule comprising a
nucleotide sequence encoding proteins or polypeptides of Pseudomonas
Conserved Effector Loc1 (CEL) and Exchangeable Effector Loc1 (EEL)
genomic sequences. CEL and EEL DNA are useful for imparting disease
resistance to a plant, by transforming a plant cell with the nucleic acid
and regenerating a transgenic plant from the transformed plant cell,
where the transgenic plant expresses a heterologous DNA molecule under
conditions effective to impart disease resistance, or by treating a plant
with an isolated protein or polypeptide, by applying the protein or
polypeptide in an isolated form or by applying a non-pathogenic bacteria
which secretes the protein or polypeptide, under conditions effective to
impart disease resistance to the treated plant. CEL and EEL proteins
are useful for causing eukaryotic cell death, by introducing a cytotoxic
pseudomonas protein into a eukaryotic cell under conditions effective to
cause cell death. CEL and EEL proteins are also useful for treating a
cancerous condition, by introducing a cytotoxic Pseudomonas protein into
cancer cells of a patient under conditions effective to cause death of
cancer cells, and thus treating the cancerous condition. The method
further involves administering a targeted DNA delivery system
comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
to the patient, where the targeted DNA delivery system delivers the
DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
is expressed in the cancer cells. The present sequence is
Pseudomonas syringae pv. tomato (Pto) DC3000 CEL DNA.
XX

```

SQ Sequence 30365 BP; 6828 A; 8938 C; 8516 G; 6080 T; 3 other;

Alignment Scores:

Alignment Scores:	4.57e-146	Length:	30365
Pred. No.:	1869.50	Matches:	381
Score:	85.19%	Conservative:	33
Percent Similarity:	85.19%	Mismatches:	71
Best Local Similarity:	78.40%	Indels:	1
Query Match:	77.67%	Gaps:	1

DB: 22

US-09-825-414-66 (1-487) x AAD20405 (1-30365)

```
QY 1 MethisileanglnSerzlaGlnGlnProGlyValAlaMetGluSerPheargThr 20
Db 23984 ATGCACATCAACCCGACGGCTCCACACCCGCTGTGACTGCGAGCGATAGCTTGGACA 23925
QY 21 AlaserApAlaSerleuAlaSerSerSerSerValArgSerValSerThrTherSercysArg 40
Db 23924 GCGTCCGACGGCTCTTGGCTCCAGCTGTGCGATGTCCAGCTCCGATCCAGCCACG 23865
QY 41 AspleuglnAlaIleThrAspTyrIleuYrSHisValPheAlaAlaHisArgPheSer 60
Db 23864 GAGATTAATGCGATTGCCGATTACCTGACAGATCATGTGTTGCTGCCATTAACCTGCCG 23805
QY 61 ValileglySerProAspGluArgAspAlaAlaAlaAlaHisAsnGluGlnIleAspAla 80
Db 23804 CCGGCGCATCGCGTGGTGGCCCAAGCTGACGTGACGTGACCAATGCGCAGATCACTGCG 23745
QY 81 LeuValIGluThrArgAlaAsnArgIleuThrSerGluGluGluThrProAlaThrIleAla 100
Db 23744 CTGATCGAGACGGCGCGCGCGCTGACCTGCATTCGAAAGGGAAACCCCGCAACCTGCC 23685
QY 101 GluThrPheAlaIleAlaValAGluAspPheArgIleuAlaThrThrAlaSerSerAlaPhe 120
Db 23684 GACACCTTCGCGCAACGCGGAAAGCTGCACCATGCGGACGACACTACATCAAGCCCGTTG 23625
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnIleuThrMetGlnProAlaIleAsnLys 140
Db 23624 CGGCGGACGCCCTTGGCCATGCGCTGCTTCAGTACATGACACCTGCGATCAACAG 23565
QY 141 GlyAspThrPleuAlaIleThrProleuYrProleuThrProleuIleSerGlyValaIleuSer 160
Db 23564 GCGCATTTGGCTGCGCGCTCCGCTCAACCGCTGACCCGCTCATTTCCGGAGCGGTGTCG 23505
QY 161 GlyAlaMetAspGlnValAGluThrLysMetMetAspArgAlaArgIleAspleuHisThr 180
Db 23504 GCGCGCATGAGACACGAGTGGGACCAAGATGATGAGCCGCGGACGGGATGTCATTAAC 23445
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 23444 CTGACCGCTGCTGCGCGACAGGCTCCACATGCGATGCGCGCTTCGTAAGGCCACACTCG 23385
QY 201 ProAlaLeuGluArgGlnValValaAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 23384 CCAGACCTTGCTCGACAGATTCGTGACACGGGGGTGGGTTCACAGTACTCGCGCGGC 23325
QY 221 AsnValValArgThrValIleuAlaProAlaLeuAlaSerArgProSerValGlnIleAla 240
Db 23324 AACGCGTACGACGATGATGGCTCGGCTCGGCTCCAGACCCGCGTGAAGGTCCT 23265
QY 241 ValAspPheGlyValSerThrAlaGlyIleuValAlaAsnAlaGlyPheGlyAspArg 260
Db 23264 GTGACCTTGTGTATCATAGCGGCTGCTGCTGCTCCACGACGCGCTTGGCAACCGC 23205
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyValAlaPheValLeuGlyMetLys 280
Db 23204 CTGCTCACTGTGCACTGCGGATACCAAGCGTGGCGGTGCTTGAAG 23145
QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaThrLysAla 300
Db 23144 GATAAAGAGCCCAAGCGTCACTAGTACGGAAGAAACGACCTGCTGAGAGCTTATAACGA 23085
QY 301 IleYrSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
```

```
Db 23084 ATCAATCGCCACCTACTCGGTGCGCGCTCAACCTGCGACGCGAATGGCGGTCTG 23025
QY 321 ProleuSpValAlaIleThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
Db 23024 CCACGTGATATGGGACCGACCGCAATGAGTGCCTGTAAGAACCTGGTGTACGCTCCAGC 22965
QY 341 LeuThrLysAsnGlyLeuAlaAlaLeuAlaGlyIleThrAlaGlyValSerLysLeuGlnLys 360
Db 22964 CTGACCCAAACGGTCTGGCCCTGCGCGGTGCTTCGAGGCGTGAAGCAAGTTCCAGAG 22905
QY 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 22904 ATGGCGACGAAATAATCATCCAGACCGCGACCGCAAGCGCGGTCAGTCACTGATCAAC 22845
QY 381 LeuValGlySerValGlyValPheAlaGlyThrThrThrThrAlaGlyLeuAlaThrAspPro 400
Db 22844 CTGCGAGGTGCGCACCGCTTTTGCAGGCTGGACCAAGCGCGCTGACACACGATCC 22785
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
Db 22784 GCGGTGAAAAAGCCGAGTCTTCATACAGACACGCGGAATGAGATGCACTCCATCAC 22725
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 22724 ACAGGCTACGTACGACGACCAAGACCTCAACTGCGCAAGACCTCAAAAGACATGGCGGG 22665
QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValaAsnLeuArgHisArg 460
Db 22664 GAGGCGATACCAATCCGCGCGCAAGCTTGGCAATACGGTCAATACCTGCGCAACGC 22605
QY 461 SerAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArgSerGluThr 480
Db 22604 CCGGCTGCTGAAGCGATATGAAGAGGGGCGACGCGGCT---TCTCCAAGTGAATA 22548
QY 481 ProPheGlnLeuArgArg 486
Db 22547 CCGTTTCGCGCTATGCGG 22530
RESULT 5
ABL29757
ID ABL29757 standard; DNA; 5397 BP.
XX ABL29757;
AC ABL29757;
XX XX
XX 26-MAR-2002 (first entry)
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40744.
XX XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX XX
OS Drosophila melanogaster.
XX XX
XX W0200171042-A2.
XX XX
XX 27-SEP-2001.
XX XX
XX 23-MAR-2001; 2001WO-US09231.
XX XX
XX 23-MAR-2000; 2000US-191637P.
XX XX
XX 11-JUL-2000; 2000US-0614150.
XX XX
XX (PEKE ) PE CORP NY.
XX XX
XX Venier JC, Adams M, Li PWD, Myers EW.
XX XX
XX WPI; 2001-656860/75.
XX XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX XX
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX XX
XX interactions -
```

PS Claim 1: SEQ ID NO 40744; 21np + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 5397 BP; 1698 A; 1618 C; 1178 G; 903 T; 0 other:

Alignment Scores:

Pred. No.:	0.019	Length:	5397
Score:	148.50	Matches:	102
Percent Similarity:	31.09%	Conservative:	64
Best Local Similarity:	19.10%	Mismatches:	231
Query Match:	6.17%	Indels:	137
De:	23	Gaps:	17

US-09-825-414-66 (1-487) x ABL29757 (1-5397)

QY 12 GYVALAlamMetCuserheargThrAlaSerAlaSerLeuAlaSerSerVal 31
DB 3 GGACAGCAGCAGCAGATAGTTACAGCACGAAATTTGTCAGCAGCAAGTTGTACGA 62
QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuShis 51
DB 63 ACAGCAGCCACAAACAAATCTGTGCCGT----- 92
QY 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
DB 93 -----TTCCGCGCAATCAT-----ACACGACATCTAATGGCGCAGCA 131
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSer 91
DB 132 ACAAAATTTTGAACAGAAATCAAAATGCAATTTGCAGCAAAATCGGCAAGACAAACAT 191
QY 92 GluGlyGluThrProAla-----ThrIleAlaGluThrPheAla 104
DB 192 GAGCGCCGCAACAGCAGCAGCAGCAACAGCAAAATTTCAACAACAGCTCACAAGTTATGCA 251
QY 105 LysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThr--- 123
DB 252 GCAGCAGCAGCAATTCAGATGCAAAATGCAGCAGCAACAAACAGCAAGATCTACACA 311
QY 124 -----ProPheAlaAlaAlaSerValLeuGlnTyrMet 134
DB 312 ACAACAGTGCCTCGGATGTCMAAACCGTTC--GCCGTAATGCCACCAAGTACGCC 368
QY 135 GlnProAlaIleAsnLysGly-----AspTyrLeuAlaThrProLeu 148
DB 369 ATCCCATGCTTGTCAGCAACAGCAGCATGATGCACTACCAATATGTTGCCGCCACAATC 428
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThr 168
DB 429 GCGCGCAACAATGCAATCCCAAGTCCCAAAATGATCGGCCACCTCGCCGCGCGCA 488
QY 169 LysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeu 188
DB 489 AAACGCCGCAACAACATTTGCCGCGCAGCAACAGCAAAATGCAACAGAAATGCACACAGA 548
QY 189 HisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGluLysArgGlnVal 208
DB 549 CATGATGGCTCGCCACAACAAC-----TCAACCAACATCTAT 587
QY 209 AspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValArg----- 224
DB 588 GATGTGCGCGCAACAGCAGCAACAGATGCAAGCCCTTTCACAGCCGCTCATCAACAGCA 647

QY 225 -----ThrValLeuAlaProAlaLeuAlaSerArgProSerVal 237
DB 648 GCGCATGCAATTCAGCAGCAACCAATTCGCCCAACAGCAGCAACACAGCAACAGCA 707
QY 238 GlnGly-----AlaValAspPheGlyValSerThrAlaGly 250
DB 708 GAGTCCCGCAGCATTTTCCGCCAATCGCCCAAGATTTCAAACCTCACCACCATGACG 767
QY 251 LeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeu 270
DB 768 AAAGTCGATCAGCATCAGTGGT-----TCTGACCAAGCCAC 806
QY 271 ArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAla----- 286
DB 807 TCAGCCGCAACATCTTCCAGCAGCAAGCTTATGCAGCAACGCCACCCGTTCCAGT 866
QY 287 -----AlaLeuSerGluGluThrAspTyrLeuAspAlaTyrLys 299
DB 867 AGCTCAAGTCAGTCAGTCATGTCAGCAATGACGCAATCAGACTCGTGTATAT----- 920
QY 300 AlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
DB 921 -----GAGCAGCAACAATTCAGCAGCAGCAACAGATGCAACAGCAACCAATGGC 974
QY 320 LeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThr 339
DB 975 ACCGCAAGCCAGCAGCAACCAATGGCAGCAGCAACCCGCAACAGCAGCAGCAGCAACCC 1034
QY 340 SerLeuThrLysAsn-----GlyLeuAlaLeuAlaGlyLysTyr--- 352
DB 1035 ACAACAGCAGCAACACCCCTTCGCGCAATCGCCACAGCAACAGCAACCTACTACCC 1094
QY 353 -----AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAla 370
DB 1095 CACCTTGCAACAGCAGCAGCAACCAACCAACGCTCAACAAATACAGCAGCAACCAAC 1154
QY 371 ThrLysAlaAlaValSerIleLeuSerAsnLeuValGlySer-----Val 385
DB 1155 ACAACAGCAGCAGCAGCAACCAACCAACAGCAGCAGCAACCAACAGCAGCTCTGACGCA 1214
QY 386 GlyValPheAlaGlyTyrThrAla-----GlyLeuAlaThrAsp 394
DB 1215 GCAACAGCCACAGCCTGCAACAGCAACAGCAACAGTCACTCAGCCGCTTATAACAC 1274
QY 395 -----GlyLeuAlaThrAsp 399
DB 1275 ATCCAGCGCTCAAGCAGCAACCAATCATTCAAAGTCATATGCTTGGCCTGCAGAACGA 1334
QY 400 ProAlaValLysAlaLeuSerPheIleGlnAspLysValLysSerThrAlaSer 419
DB 1335 GCAGCAAT-----GTTGACGGTTCACCAACCAACCAAC 1367
QY 420 ThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSer 439
DB 1368 ACAGCAGCCGCAACAGCAACAGATTTACTCAAGTGCAGCAACTGCCACCGCACAGCA 1427
QY 440 GlyLysAlaIleSerSerThrGlyAlaSerLeuArgSerThr 453
DB 1428 GCAGCAACAGCTTCCGCAACAGCAGCAAGTTCAACAACAGCA 1469

RESULT 6
AAD31881 standard; DNA; 4645 BP.
ID AAD31881;
AC AAD31881;
XX 18-JUN-2002 (first entry)
DT Lactobacillus rhamnosus outer membrane protein rompa' gene.
DE Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;
KW fermentation process; anti-infection; rotavirus infection; heart disease;

KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW rompa gene; ds.
XX
OS Lactobacillus rhamnosus HN001.
XX
FH Key Location/Qualifiers
FT CDS 74..4465
FT /*tag= a
FT /product= "Outer membrane protein"
FT /note= "CDS does not include start codon"
FT /partial
XX
XX MO200212506-A1.
XX
XX 14-FEB-2002.
XX
XX 08-AUG-2001; 2001WO-NZ00160.
XX
XX 08-AUG-2000; 2000US-0634238.
XX
XX 28-NOV-2000; 2000US-0724623.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX PA (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.
XX
XX Glenn M, Havukkala IJ, Bloksberg LN, Lubbers MW, Dekker J,
XX Christensson AC, Holland R, O'toole PW, Reid JR, Coolbear T;
XX WPI: 2002-241760/29.
XX P-PSDB; AAE20110.
XX
XX New polynucleotides and polypeptides from Lactobacillus rhamnosus,
XX useful in e.g. improving the flavor, aroma, texture and health-related
XX benefits of milk-derived products, or in increasing properties of
XX microbes -
XX
XX Claim 2: Fig 63; 257pp; English.
XX
XX The present invention relates to a new isolated polynucleotide comprising
XX a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a
XX polypeptide capable of modifying the flavour, aroma, texture, nutritional
XX and health benefits of milk-derived products, and/or survivability of
XX microbes in dairy manufacturing processes. The polynucleotides are useful
XX for improving the properties of microbes used in the manufacture of milk-
XX derived products such as cheeses, yogurt, fermented milk products, sour
XX milks and buttermilk; in modifying the flavour, aroma, texture and health
XX -related benefits of milk-derived products and in increasing the survival
XX of microbes during industrial fermentation processes. The bacteria may be
XX used to increase resistance to enteric pathogens and anti-infection
XX activity, including treatment of rotavirus infection and infantile
XX diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;
XX liver cancer reduction; reduction of small bowel bacterial overgrowth;
XX immune system modulation and treatment of autoimmune disorders and
XX allergies; treatment of allergic responses to foods; reduction of blood
XX lipids and prevention of heart disease; antihypertensive effect;
XX prevention and treatment of urogenital infections; Helicobacter pylori,
XX or hepatic encephalopathy; treatment of inflammatory bowel disorder and
XX irritable bowel syndrome; modulation of endocarditis; and for improved
XX protein and carbohydrate utilization and conversion. The transgenic
XX microbial population can be administered to a mammal as an anti-
XX carcinogenic agent. The present sequence is Lactobacillus rhamnosus
XX outer membrane protein rompa gene.
XX
XX Sequence 4645 BP; 1279 A; 1062 C; 1261 G; 1043 T; 0 other;
XX
XX Alignment Scores:
XX Pred. NO.: 0.021 Length: 4645
XX Score: 147.00 Matches: 104
XX Percent Similarity: 33.53% Conservative: 64
XX Best Local Similarity: 20.76% Mismatches: 241
XX Query Match: 6.11% Indels: 92

DB: 24 Gaps: 12
US-09-825-414-66 (1-487) x AAD31881 (1-4645)
QY 1 MethisIleasnGlnSerIalGlnGlnProProGluValAlaMetGluSerPheArgThr 20
DB 2213 ATCAACGCCAATAGTGTGCTGATGTCACAGCTTCACAGCAAGGCGTACAGTCTGCT 2272
QY 21 AlaserAspAlaserLeuAlaSerSerValArgSerValSerThrThrThrThrThrThr 40
DB 2273 GCTAGTGATGCGGCGAGTATTATGCAAGTCAAGCCAGATGATGCTGCGAGCATGCTGAT 2332
QY 41 AspleuGlnAlaIleThrAspTyrLeuYshHisValPheAlaAlaHisArgPheSer 60
DB 2333 AATATGGA-----ATCAAGTCTCTCGCCAGTATGCTGAGAGCAATCG 2377
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisasnGluGlnIleAspAla 80
DB 2378 CAAATTCCTTGGCAGCTAGCAAGTCTGCTGCGCTAGTCCAGTCCGCGAGCGTCCGCA 2437
QY 81 LeuValGluThrArgAlaAsnArgLeuYSerSerGluGluGluThrProAlaThrIle--- 99
DB 2438 GCAATCGTG-----GCAAGTAGCGCGGCTAGTAGAGCGTCACTGCAAGCTGCGCTA 2491
QY 100 -----AlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
DB 2492 AGTAAGCGTGATGCATGATGCAAAAGCTCTGACCGGCTGCTATGATTCCTAGCGTTCTGAG 2551
QY 116 AlaserSerAla-----PheGluAsnThrProPheAlaAla 127
DB 2552 GCCAGTGGCGGCTTCGCTGCTGCAATGATGATTCGGGATATGCGACATGATTCGACACA 2611
QY 128 AlaserValLeuGlnThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 147
DB 2612 AGTTCGCGTGGGCGGCGCATG----- 2632
QY 148 LeuYsProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 167
DB 2633 -----AGCGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2668
QY 168 ThrYsMetMetAspArgAlaArgGlyAspLeuHisThrIleSerThrSerProAspArg 187
DB 2669 GTTGCAAGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2698
QY 188 LeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnVal 207
DB 2699 -----ACTGCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2752
QY 208 ValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArg-----Thr 225
DB 2753 ACTGCAGCTAGCAGCAAGCAAGTCAAGCTGATGATGATGATGATGATGATGATGATGATGAT 2812
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly----- 239
DB 2813 GATTACGCAAGTGGCGCAAGCTCCAGTCCAGCAAGCGGGGTCAGATGCACTGCAACA 2872
QY 240 -----AlaValAspPheGlyValSerThrAlaValGlyLeuValAlaAsn 254
DB 2873 TCTGCGTATGCTAGTGCAGCAAGTTCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 2932
QY 235 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274
DB 2933 GCAGGCTCCAGGCGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 2992
QY 275 PheValIleuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTrp 294
DB 2993 ACCGCTGCAGTGCAGGCTCCAGTCCAGTATCCGAAGATATGCGGATTCAGTCACTAGCCAGT 3052
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
DB 3053 CAGGCTGCAGGCGAGGCGCAAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3109
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSer 334

```

Db 3110 -----GGGCGCGTTGTTTCAAGTGTGCGCAGTATGCAAGTGAACAGCGCAAGACG 3160
Oy 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyThrAlaGly 354
Db 3161 GCTGCACATGCGGATGTGGTGCACAGAGTGGCGCCAGCCAGCAGTAACTAATCGAGT 3220
Oy 365 ValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla 374
Db 3221 GCGGCAGCC-----AGTGCAGCAAGAGCTGGT 3247
Oy 375 ValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrThrAla 394
Db 3248 GATACCAAGCCGCA-----GCAAGATCTCGAGTGCA 3280
Oy 395 GlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLys 414
Db 3281 GCGAGTGTGTCAGCAAGCAGTCCCAAGGTGTCAGAGCAGTGTCCAGCAGCGCGAGT 3340
Oy 415 SerThrAlaSer-----SerThrThrSerTyrValAlaAspGlnThrValLysLeuAla 432
Db 3341 GCGCGCGCATCCGATGACTCGTACCTTCTAGTCCGCGCAGTGGCGCTGCAGGCTTTGAC 3400
Oy 433 LysThrValLysAspMetSerGlyGlnAlaIleSerSerThrGlyAlaSerLeuArgSer 452
Db 3401 AAGGTGCACAGCGCTGGCGAAGCGCAGCTTCAAGTGCAGCGCGCGCTCTAGTTCA 3460
Oy 453 ThrValAsnAsnLeuArgHisArgSerAlaProGlnAlaAspIleGlnGlyGlyLe 472
Db 3461 GCGCAGAGTCAAGGACACAGAGTGGCGCAAGCTCAGTGCAGCAGGAAGCGGTCAAGCA 3520
Oy 473 Ser 473
Db 3521 TCA 3523

RESULT 7
ABL23378/c
ID ABL23378 standard; DNA: 3146 BP.
XX
AC ABL23378;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 21607.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PERK ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI: 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX
PT interactions -
XX
PS Claim 1: SEQ ID NO 21607; 21pp + Sequence Listing: English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
XX
CC capable of detecting 1000 or more genes from Drosophila. The invention is
```

```

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL161716-ABL35111), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABR57737-ABR72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcr_sequences.
XX
SQ Sequence 3146 BP; 783 A; 811 C; 727 G; 825 T; 0 other:
XX
Alignment Scores:
Pred. No.: 0.0141 Length: 3146
Score: 146.50 Matches: 91
Percent Similarity: 39.32% Conservative: 60
Best Local Similarity: 23.70% Mismatches: 156
Query Match: 6.09% Indels: 77
DB: Gaps: 13
XX
US-09-825-414-66 (1-487) x ABL23378 (1-3146)
Oy 141 GlyAspTyrPheAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 225 GCGCGCTTAGTGGCGCGAGGCTTCAGACGAGCAGGAGCTTGTAGCGGTACCATCGCT 2196
Oy 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis 179
Db 2195 GCGCAGAGCGGTCTGTGCGCGCACATCATATGCGCAGAGGCGTCTCTCCACACGG 2136
Oy 180 -----TyrLeuSerThrSerProAspLysLeu 188
Db 2135 GCGCAGTACTCGCCGGAAGCAGCTCAGACATCAGCGCGCGAGATCAGACAGCAG 2076
Oy 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGly 204
Db 2075 CACGGTCTTGTAGCGGTATCCATCATCAGCCAGGGGAGTCTCAGACGACAGCCTCGGA 2016
Oy 205 ArgGlnValAlaAspMetGlyIleAlaValGln-----ThrPheSer 218
Db 2015 GCGGAGCTCTCTACACAGGGGCGAGTATTCAGAGGAGCAGACTCGAGACATCAGCAGCAG 1956
Oy 219 AlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238
Db 1955 ACGGATCACACGACGACGCGTCTGTAGCGGTATCATCATCAGCAGGAGTCTCGAC 1896
Oy 239 GlyAlaValAlaAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGly 258
Db 1895 GGC-----AACGGCTTCGGA 1881
Oy 259 AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGly 278
Db 1880 GCGGATTCCTC-----AACAGGGGCGAGTATTCGCGGACAGCTCGCTTAC--- 1833
Oy 279 MetLysAspLysGlnProLysAlaAlaLeuSerGlnGluThrAspTyrLeu----- 295
Db 1832 -----ATCACAGCGCGAGCGATTCACACGACGACGACGACGATCAGCGG 1788
Oy 296 -----AspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
Db 1787 TCCATCAGCAGCAGCAGGGGCTTCAACAGCAGCAGCTCGGAGCACTCTCAACAGGG 1728
Oy 314 GlyLysArgMetAlaGlyLeuProLeu---AspValAlaThrAspGlyLeuLysAlaVal 332
Db 1727 GCGCAGGTACTCGCCGGAAGAGCTCATTTGACATCAGCAGCAGCAGCAGATCAGCGGG 1668
Oy 333 ArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu-----AlaLeuAla 349
Db 1667 GACGCTCTTGTACAGGTATCATCATCAGCAGAGGGGTGTCCACAGCAGACGCTCAGA 1608
Oy 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
Db 1607 GCGGACATCTTCACAGGCGGAGGTATTCAGGGGACAGCTCGAATCATCAGCAGCAGCAG 1548
```

QY 370 AlArthrlyAlaValSerGlnLeuSerAsnLeuValGlySerValGly---ValPhe 388
||| ||| : : : : : ||| ||| : : :
Db 1547 ACGGATACACAGCGGCGACAGTCTTGACCGTACCCATCATCAGCCAGGAGCTCTGCAC 1488
QY 389 AlaGlyTrpThrAlaGlyLeuAlaThrAspProAlaValLysAlaGluSerPhe 408
: : : : : ||| ||| : : : : :
Db 1487 AGCAACAGCCTCGGAGGCGACTCTCCACAGGGGCGAGTACTCAGGGGAGAGTTC--- 1431
QY 409 IlleGlnAspLysValLysSerThrAlaSerSerThrThrSerThrValAlaAspGlnThr 428
1430 -----GGAGACATCACGACGACGAGGATCACACGCGGAGGCTCTT 1389
QY 429 ValLysLeuAla-LysThrValLysAspMetSer-----GlyGluAlaIleSe 444
||| ||| : : : : : ||| ||| |||
Db 1388 GTACAGGTAGCGCATCATCAGCCAGAGAGTCTCGACGCGCAACGGCTTCGAGGCGGATTC 1329
QY 444 rSerThr-GlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProG 464
||| ||| ||| ||| : : : ||| ||| |||
Db 1328 CTCACACAGGGGCGAGGTATP-----CACCGGAGGGGAGCTC 1293
QY 464 IuAlaAspIleGluGluGlyIleSerAlaPheSerArgSerGluThrProPheGlnL 484
1292 GTTACATTCACGACGCTGGCGCAACTTACGCGGCGACGCTTGTAGCGGATTCATC 1233
QY 484 euArgArg 486
||| ||| ||| ||| ||| |||
Db 1232 CTCGCGAG 1225
RESULT 8
AAD04029 standard; DNA; 9542 BP.
ID AAD04029;
XX AAD04029;
XX 02-JUL-2001 (first entry)
XX Moraxella catarrhalis outer membrane protein-106 (OMP106) DNA.
DE Moraxella catarrhalis outer membrane protein-106 (OMP106) DNA.
XX Outer membrane protein-106; OMP106; haemagglutinating; vaccine;
KM bacterial infection; immunogen; cytotoxic; antibiotic;
KM passive immunisation; ds.
XX Moraxella catarrhalis.
OS Moraxella catarrhalis.
XX
FH Key Location/Qualifiers
FT CDS 218..6589
FT /*tag= a
FT /product= "Outer membrane protein-106 (OMP106)"
PN US6214981-B1.
XX 10-APR-2001.
PD 10-APR-2001.
XX 12-NOV-1997; 97US-0968685.
PE 03-MAY-1996; 96US-0642712.
XX (ANTE-) ANTEX BIOLOGICS INC.
PA Tucker K, Plosila L, Tillman WF;
XX
PI WPI: 2001-281002/29.
DR P-PSDB: AAE00701.
XX
XX Novel nucleotide sequences encoding Moraxella catarrhalis outer
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
PT infections and as vaccine against Moraxella catarrhalis infection of
PT mammals -
XX
PS Claim 6; Column 45-54; 49pp; English.
XX
CC The present sequence is haemagglutinating Moraxella catarrhalis outer

CC membrane protein-106 (OMP106) DNA. The OMP106 is used as a therapeutic
CC and prophylactic vaccine against M. catarrhalis infections of mammals.
CC It is used for diagnosis of bacterial infections and as reagents for
CC clinical or medical diagnosis of M. catarrhalis infections and for
CC scientific research on the properties of pathogenicity, virulence and
CC infectivity of M. catarrhalis. It is also used as a probe to identify
CC the presence of M. catarrhalis in biological specimens and to identify
CC other bacteria that encode a polypeptide related to M. catarrhalis
CC OMP106. OMP106-derived polypeptides are used as ligands to detect
CC antibodies elicited in response to M. catarrhalis infections and also
CC as immunogens for inducing M. catarrhalis-specific antibodies which are
CC useful in immunoassays to detect M. catarrhalis in biological specimens.
CC Cytotoxic antibodies are useful in passive immunisations against
CC M. catarrhalis.
XX
XX Sequence 9542 BP; 2966 A; 2104 C; 2098 G; 2374 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0578 Length: 9542
Score: 146.50 Matches: 111
Percent Similarity: 36.00% Conservative: 69
Best Local Similarity: 22.20% Mismatches: 181
Query Match: 6.09% Indels: 139
DB: 22 Gaps: 23
US-09-825-414-66 (1-487) x AAD04029 (1-9542)
QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHis 51
: : : : : ||| ||| ||| ||| ||| |||
Db 1346 AAGACTTAACTATACACAGGGGTGCACAGACGCGCATTAACCGAT----- 1393
QY 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
||| ||| ||| ||| : : : ||| ||| |||
Db 1394 -----CATACATCGTGCTGTGATCAAAATGGCGCTGCAAAATTCAA 1438
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArg-----AlaAsnArg 88
||| ||| ||| ||| : : : ||| ||| |||
Db 1439 CTTCCTGAACCTTAAACCGCTTAAATGGTATGCCCTGAAACCTTAACCGGACAGAG 1498
QY 89 LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys 108
: : : : : ||| ||| ||| ||| ||| |||
Db 1499 AAAGTTACCGTAGGCAAAAC----- 1519
QY 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla 128
||| ||| ||| ||| ||| ||| |||
Db 1520 -----CGCCTT---ACCACAGATAAATGCTTTACCAATGAT-----ATG 1558
QY 129 SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeu 148
: : : : : ||| ||| ||| ||| ||| |||
Db 1559 AATGGCATTTGATGAAGCAAAACCTTATCTTGATGAAGACACTGGCATTCATGCAGTGGT 1618
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeu----- 159
Db 1619 CAAGAATTCACCAACTTACTGCTGGTGTAGTACATGCAGTGCAGACTTATGACAG 1678
QY 160 -----SerGlyAlaMetAspGlnValGlyThrLysMetMet 171
Db 1679 CTTAAAAAAGTTAACCAAAACCGCTGAAGTGCCTCTACAAACCTTACCGTTAAAAAGTGA 1738
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
||| ||| ||| ||| : : : ||| ||| |||
Db 1739 GATAAATAGTGAATGAT-----GCTAATGACAGC 1768
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMet--- 210
Db 1769 AAAATCATCACCGTGGTAAAAATACCAAAACCAACGCTACTCAAGTCAACACCTTAAA 1828
QY 211 -----GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
||| ||| ||| ||| ||| ||| |||
Db 1829 CTCAAAGGTGAACACGGTGTGATGTACAAAC----- 1861
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal 245
||| ||| ||| ||| ||| ||| |||

Db	1862	-----GAAACAANAATGTCACGTTACCTTGGCGTT	1891
Qy	246	SerThralaGlyLeu---ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerAla	264
Db	1892	ACCCAAATACCGCTCTGACCGCTGGACACAGACCCTTAACACAGATGGCTTATCTGT	1951
Qy	265	GLN-----SerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLeuAsp	281
Db	1952	AAAACACCAATGTATACAAACAATTCCAAGTCGGGGTGATGCAATTACATTACTAT	2011
Qy	282	-----LysGluIrrProLysAlaAlaLeuSerGluIuThr-----AspTrr	294
Db	2012	ATCAGCAATAGTAAGCAGAGTGCTGGCATTGAAANAATACACCTCGCATTAACGAGAGCGT	2071
Qy	295	LeuAspAlaTyrItyAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly	314
Db	2072	ATT-----GGTTTTCCTAAATATACTGCT	2095
Qy	315	LysArgMetAlaGlyLeuProLeuAspValAlaThrAsp-----GlyLeuLysAla	331
Db	2096	TCATTGGATGCAACAACACCCCGCTTAACCCCACTGGCATTAACGAGGTGTAAAG	2155
Qy	332	ValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly	351
Db	2156	CTGACCAATGTCCAAATCTGCATTAACCTCGTACCAATGCT-----	2197
Qy	352	TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThr	371
Db	2198	---GGCAGCTTAACCTTTATGAACCGCCCTTAACGACTGCTAAATACCGAANAATCAGGCTCT	2254
Qy	372	LysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyr	391
Db	2255	GCCGCCACCACTTAAGACCTATACACCTATACACAAGTACGCGCTACCTTGGACGT---	2311
Qy	392	ThrThrAlaGlyLeuAlaThrAspProAlaVal---LysLysAlaGluSerPheIleGln	410
Db	2312	-----GATACAGGCTCCTAATGTGCACCAAAAACCTGGCGAGATTTTAAG	2356
Qy	411	AspLys---ValLysSerThrAlaSerSerThrThr-----SerTyrValAla	425
Db	2357	GTTAAAGCTGTAAACACACAGCTGATGATTTAACCAAAAATTAACATGGGTGTGGCT	2416
Qy	426	AspGln-----ThrValLysLeuAlaLysThrValLysAspMetSerGly	440
Db	2417	GATGTACCGATATAGCTTAACCGTTAACTTGCTAAACCTTTAAGCGATCTGATCG	2476
Qy	441	GluAlaIleSerSerThrGlyAlaSerLeuAspSerThrValAsnAsnLeuArgHisArg	460
Db	2477	GTTAATACTTAAACCCCTAACCTGCACAGTAAGAATTACCGTAGACAGTGGCAACAC---	2533
Qy	461	SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSerGluThr	480
Db	2534	-----ACCGCTAAGCTACAAATAATGGTGAATTAAAC---TTTAGCAACAATAATCA	2581
RESULT 9			
ABL29756/c			
ID	ABL29756	standard; DNA; 7029 BP.	
XX	AC	ABL29756;	
XX	DT	26-MAR-2002 (first entry)	
DE	XX	Drosophila melanogaster genomic polynucleotide SEQ ID NO 40741.	
KM	XX	Drosophila; developmental biology; cell signalling; insecticide;	
OS	XX	pharmaceutical; gene; ds.	
XX	OS	Drosophila melanogaster.	
XX	FN	W0200171042-A2.	
XX	XX	27-SEP-2001.	
XX	XX		

XX	23-MAR-2001; 2001WO-US09231.
XX	
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	(PEKE) PE CORP NY.
XX	
P1	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI; 2001-656860/75.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
XX	interactions -
XX	
PS	Claim 1; SEQ ID NO 40741; 21np + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB101840-AB116175) and the encoded proteins
CC	sequences (AB101840-AB116175) and the encoded proteins
CC	(AB116173-AB116175).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 7029 BP; 1426 A; 1440 C; 1945 G; 2218 T; 0 other;

Alignment Scores:			
Pred. No.:	0.0934	Length:	7029
Score:	142.00	Matches:	112
Percent Similarity:	31.338	Conservative:	65
Best Local Similarity:	19.82%	Mismatches:	240
Query Match:	5.90%	Indels:	149
DB:	23	Gaps:	18

US-09-825-414-66 (1-487) x ABL29756 (1-7029)

QY	2	HislleasngInserlaIngaIngInProprogiUvalAlametGluSerPheargThrala	21
Db	6820	CACCACACAGCAGATAGTTCMAACAGCA-ACAGGTTGGCAGCAACAGTGGCCAGCAGCA	6762
QY	22	SetrpaSaSerleuAla-----SerSerSevalaIngSerValSerThThSerGys	39
Db	6761	AACCATGGTACACAGCAGCAGATAGTTCAGCAACAGCAATGTTTTCAGCAGCAGATTGT	6702
QY	40	ArgsrpreuInglaIleThrasrUyleuLysHisHisValPheAlaAlnIaIarPhe	59
Db	6701	TCAGCAACAGCAGCCAGCCACACACATCTGCCGGT-----TTGCCGCAACATAT---	6651
QY	60	serValleGlySerProAsrGluIarGsrAlaAlaIeAlaAlnIaHisngInglInleAsr	79
Db	6650	-----ACACCAACATCTAATGGCGCAGCAACAAATTTTGACACAGATTCACAT	6603
QY	80	AlAleuValGluThraIarGlaAsnIarGleUyrSerGluGlyGluThrProAla-----	97
Db	6602	GCAATTCACAGCAATCTGTCMAAGCCGACACAGCATTTGACGCCGCAACAGCAGCAGCA	6543
QY	98	-----ThrIleAlaGluThrPheAlaLysAlaGluLysrPheAsrIarGleu	112
Db	6542	GAGCAAAATTCACACACACAGCTGCACACAGATTTGACGACGACGACGACATTCACAGATGCA	6483
QY	113	AlaThrThralaSerSerAlaPheGluAsnThr-----	123
Db	6482	AATTCACAGCAGCAACACAGCAAGATTCACACACAAACAGTCGCGCTGGATGTGCA	6423
QY	124	---ProPheAlaAlaAlaSerValleuGlnUyrMetGlnProAlaIleAsnLysGly---	141
Db	6422	AACACGTTTC---GCCGGTATTCACACAGTATGCGCATTCGCGCATTCGCGCAGCAACAGCA	6366

Qy	142	-----AsprtleuAlaThrProLeuLysProLeuThrProLeuIleSer	156
Db	6365	GCATCATGCACTAGCAATATGTGGTGGCCACAAATCGCCGACAACTGCATGCCCA	6306
Qy	157	GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetIleAspArgAlaArgGly	176
Db	6305	TTCCCAATGACGTCCGCCACCTCCGCCCTCGCCGCAAAAGCCGACAAACAATTCG	6246
Qy	177	AspLeuHisThrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerVal	196
Db	6245	CCAGCAACAGCAAAATGCACAGAAATGCAGAACAGCAACGACATGATGGGCTCGCCACAACACC	6186
Qy	197	LysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyTyrIleValGlnThr	216
Db	6185	-----TCMACCAAATCTATGATGTGGCGCAACAGACAGCAACA	6147
Qy	217	PheserAlaLeuAsnValValArg-----	Thr 225
Db	6146	GATGACAGCCCTTTACACAGCCCGTTCATCAACAGCAGCCCATGCAATTGCACAGCAACA	6087
Qy	226	ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly-----	239
Db	6086	ACAATGTGCGCCACAGCAGCAACACAGCAACAGCAGAGATCCCGACATATTTCGCCGA	6027
Qy	240	---AlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGly	258
Db	6026	ATCCGCCCAAGATTCTCAAACTCCACCCCAATGCAGGCAAAAGCTCAATCAGCATCAGTGGT	5967
Qy	259	AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyValAlaPheValLeuGly	278
Db	5966	-----TCCTGGACAGCCACTCAGCCGACATCTATTCTCCA	5928
Qy	279	MetLysAspLysGlnProLysAla-----	Ala 287
Db	5927	GCAGAACGCTATGCAGCCACAGCAGACCCCGTCAAGTAGCTCAAGTGCATGCATCTGC	5868
Qy	288	LeuSerGlnGluThrAspThrPheAsnAspAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db	5867	ATTGACAGCAATCTGAGCACTCGTTGATTAT-----GAGCGACAAACATGTGA	5820
Qy	308	GlyAlaIleAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAsp	327
Db	5819	GCACCAACAACAGATGCGACACAGCAACAACAATGCGACCGCAGCCCGACAGCACACAAT	5760
Qy	328	GlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla	347
Db	5759	GGCAGACGACACCGCAACAGCAGACGACGACGACACCAACAAGCAGCACACCCCTTCTCC	5700
Qy	348	LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThr-----	363
Db	5699	GCGCAATTCGCCACTGCAACAGCAACCACTACTACCCCGACCTTGCAACAGCAGCAACCA	5640
Qy	364	-----LysAsnIle-----	366
Db	5639	ACAAACGCTCAAGTAAATATAAGAACATATATATTTTACCGGAAATTAATGATTAC	5580
Qy	367	-----ThrAspSerAlaThrLysAlaAlaValSerGlnLeu	378
Db	5579	TTCTTTTACCAATATACAGCAGCAACAACAACAACAACAGCAGCAGCAGCAACAACA	5520
Qy	379	SerAsnLeuValGlySerValGlyValPheAla-----	GlyTyrThrAla 394
Db	5519	ACAGCAGCAGCAACAACAGCAGGTCTCTGACGACGACAAAGCCACAGCCTGGACAAACAGA	5460
Qy	394	-----	394
Db	5459	ACAAGTCATCACTCAGCGCATGTTATAACAACATCCAGCGCTCAAGCAACAACATCAT	5400
Qy	395	-----GlyLeuAlaThrAspProAlaValLysValGlyAlaGluSerPhe	408
Db	5393	TCAAAGTCATATGCTTTGGCTTGGCAAGAAGCAGCAGCAAT-----	5358
Qy	409	IleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThr	428

Db	5357	-----GTTGCACGCTTCAACACAGCAGCAGCAGCCGCGACGACACACACAGAT	5307
Qy	429	VallysLeuAlaIysrThrVallysAspMetSerGlyGluAlaIleSerSerThrGlyAla	448
Db	5306	TACTCAAGTGCAGCAACTGGCCAGCGACAGCAGCAAGCAACAGCTTCGCAACAGCACCA	5247
Qy	449	SerLeuArgSerThr	453
Db	5246	AGTTCACACACACA	5232
Result 10	AAAX24980		
ID	AAAX24980	standard; DNA: 7100 BP.	
AC	AAAX24980;		
XX	05-JUL-1999	(first entry)	
DE	Bacillus subtilis metalloprotease YOMI DNA.		
XX	YOMI; metalloprotease; protease; textile; animal feed; detergent;		
KW	Gram-positive bacterium; ds.		
XX	Bacillus subtilis.		
OS			
XX	Key	Location/Qualifiers	
FX	CDS	112..6969	
FT		/*tag= a	
XX	W09J14342-A1.		
PN			
PD	25-MAR-1999.		
XX			
PE	08-SEP-1998;	98MO-US18828.	
PR	15-SEP-1997;	97GB-0019636.	
XX			
PA	(GENV) GENENCOR INT BV.		
PA	(GENV) GENENCOR INT INC.		
PI	Estell DA;		
XX			
DR	WPI; 1999-229541/19.		
DR	P-RSDB; AAM98149.		
XX			
PT	Compositions containing metalloprotease from Gram-positive microorganism		
PS	Claim 25; Flg 1A-O; 59pp; English.		
XX			
CC	This DNA sequence encodes metalloprotease YOMI (see AAM98149) of Bacillus subtilis. The invention relates to the discovery of this previously unknown metalloprotease, uses of the metalloprotease in industrial and agricultural applications, and advantageous strain improvements based on genetically engineering a Gram-positive microorganism to delete, underexpress or overexpress the enzyme. Due to overall relatedness of the enzyme with pseudomonas lasa protein, YOMI appears to be a member of the M23 metalloprotease family. The metalloprotease can be used in claimed cleaning compositions, animal feed and compositions for the treatment of textiles. It may also be used for peptide hydrolysis, waste treatment and for cleaving recombinant fusion proteins. Nucleic acid sequences are useful as primers and probes for detecting similar sequences in other organisms, and for recombinant production of the enzyme. Gram-positive cells in which YOMI is inactivated (by gene mutation or deletion) are used for production of heterologous proteins, especially enzymes, hormones, growth factors and cytokines.		
QO	Sequence 7100 BP; 2649 A; 1294 C; 1353 G; 1804 T; 0 other;		
Alignment Scores:			

Pred. No.:	0.139	Length:	7100
Score:	140.00	Matches:	128
Percent Similarity:	33.44%	Conservative:	81
Best local Similarity:	20.48%	Mismatches:	222
Query Match:	5.82%	Indels:	194
DB:	20	Gaps:	26

US-09-825-414-66 (1-487) x AAX24980 (1-7100)

Oy	15	MetC	UsherPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValAlaArgSerVal	34
Db	862	ATTCAATATTATATGACGACGACGACAGCAAGTAAATGTTCAAAACCTTAATATACAGGATATGC	921
Oy	35	SerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuIleuShiHisValPhe	54
Db	922	AGTTCTATGGGCTCTAGTATATACAGACGTTCAAGATATTATTTGAAT	969
Oy	55	AlaAlaHisArgPheSerVal---IleGlySerProAspGluArgAspAlaAlaLeuAla	73
Db	970	GCACTAAATAGTCTTATATGTAAACACTGGMAAGCAATAAATATGATACATCAAAATTCAAAGC	1029
Oy	74	HisnGluGlnIleAspAlaLeu-----ValGluThrArgAlaAsnArgLeuArg	90
Db	1030	TTGAAATTCGCAATTTTACAGATTAAGCTCCAAACGCTCAAAACAGCTATATCAACGCTCT	1089
Oy	91	SerGluGlyLeuThrProAlaThrIleAlaGluThrPhe-----	103
Db	1090	TCCTTTTGGT-----GCAGAACTTAACCCAAACCTTCAAAAGCATGTCCACCTATTTA	1140
Oy	104	-----AlaIysAla	106
Db	1141	ATCTCCGGTTCTTATATCTACGAGCATCTGTGGACTTAAAGAAATGGTATCCACGACG	1200
Oy	107	GluIysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla	126
Db	1201	ATGAAATTTGATCTCTCATGACAAATTAATTCGCGCTGTATGAAATGACGGGATTAATTA	1260
Oy	127	AlaAlaSerValLeuGlnIleTyrMetGlnProAlaIleAsnIleGlyAspTyrPheAlaThr	146
Db	1261	TATAAATGAACTTGC-----CAAGATCATTTGACTTAAAGTGATACATTTCAAT	1311
Oy	147	ProLeuIysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal	166
Db	1312	AAAATTCACAGATTTCTTCAAAATGACAGCGGATTTGGGAGAAATGGGTTTCATGAAGT	1372
Oy	167	GlyThrIleuMetLeuAspArgAlaArgGlyAspLeuIleHisTyrLeuSer---ThrSerPro	185
Db	1372	GAGCTCTCCACAGTTAAACGAAACCTGCCCAAGTTCTCAAAATGTCTGTATTAATCTCC	1431
Oy	186	AspIysLeuHisAspAlaMetAlaValSerValIysArgHisSerProAlaLeuGlnArg	205
Db	1432	GATGATACAGTTAAACACTCTAAGGCGACGACACTGCTCAACTTTAATATTCACCAAAATGAT	1491
Oy	206	GlnVal-----ValAspMetGlyIleAlaValGlnThrPhe	217
Db	1492	TCAATATCAATTCGACGATTAATTAATGAAGGTGATTAATCAATCTGTCTACAACTGTA	1551
Oy	218	SerAlaLeuAsnValValaGlyThrValLeuAlaProAlaLeuAlaSerArgProSerVal	237
Db	1552	GATCTGGCCCAATCTTATCCCTTAA-----GCGGTTCACAGCTTTTCAATTCGGGGTA	1605
Oy	238	GlnGlnIleAlaValAspPheGlyValSerThrAla-----Gly	249
Db	1606	GAGCTAATGATCTTATGTGTTATACACTGCAATTCGTAAGTACAAACAGCTGAATACAGGG	1666
Oy	250	GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer	263
Db	1666	AATATGTCGGGGAACCTCTTAAGACAAATTTTGGCGGATTTGGGAATTAATCAAGCTCA	1722
Oy	264	ValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlnIleMetIysAspIysGlu	283
Db	1726	ATTTAACCGTTAGAACAGATTTGGTATCTCAGTTTAAACAGCTGTGGTGTCAACCTTAATCA	1785

QY	284	ProlysAlaAlaLeuSerGlu-----GluThrAspTrpLeu---AspAlaTrpLys	299
Db	1786	GCAAGTATTTAAATTAAGTGAAGTTCGTGGTAAGTGGAGATACCTTTTCGATGCTCAGAAA	1845
QY	300	AlaIleLysSerIaAserTrpSerGly-----	308
Db	1846	CAAAATACCTCAATTTGGAGTAGTGGTGGATTAATTCATTAATATCCGTTTAATGCAATGATG	1905
QY	309	-----AlaIleAsnAlaGlyLys-----	315
Db	1906	AACAACTTCTCTATTCGTCAAGAAATGGCGCTAAACACGCGGCTACTCAACAGGAAGTCT	1965
QY	315	-----	315
Db	1966	TGAGTAGCAGCAGCAAAAGTATGCAGATAGCTACAGACTAGGATTAATTAAGCTCAAAAT	2025
QY	315	-----	315
Db	2026	AACCTTCACTGAATTTGCTATTTGCAGCTCTGTATGCTTTTATTAGCAGCAGATTAATGAA	2085
QY	316	-----ArgMetaIleGlyLeuProLeuAspValaIaThrAspGlyLeuLysAlaVal---	332
Db	2086	TTTACTCAACCCGCGAGGTTCTTTGCTTAAAGCTTCTACAGGATATCAAAACATTCGG	2145
QY	333	-----ArgSerLeuValSerIaThrSer-----LeuThrLysAsnGly	345
Db	2146	TTTCTACTCCCTCCCTTTTACGTCAGATAGACACTGCACCCCTTTTCTCAGTAAATAC	2205
QY	346	LeuAlaLeuAlaGly-----GlyTrpAlaGlyValSerLysLeuGlnLysMet	361
Db	2206	CGCACAATTACCCAGCAGCACCCTAATTTGGGCGACACGCTGAATGGGGCAA---GAAACTTAA	2262
QY	362	AlaThrLysAsnIleThrIraSerIaThrLysAlaAlaValSerGln-----	377
Db	2263	CGCACTGCTGGCGTACAGAGCTGGTATGACTCGTGCAGCAGCTGCCCTCAAGACTTAAAA	2322
QY	378	-----LeuSerAsnLeuValGlySerValGlyValPheAlaGly	390
Db	2323	ACTGCTTTCGAGGGTTCCTGTTTCAACTTTTAGTTGGCGGTGCATTTGCTCTTGGGA	2382
QY	391	TrpThrIraIaGlyLeuAlaThrIraProAla---ValLysLysAla-----	405
Db	2383	TGGGCGCTAGATCAATTAATTTCTCTTTTGCAGAAAGCTAAAGACCTAAAGATGATTTT	2442
QY	406	-----GluSer-----	407
Db	2443	GAGCAGAGCCAGCAAAACCAATGTGCAAGCAATTAGCACCATAAAGACTCCACTATATAA	2502
QY	408	PheIleGln-----AspLysValLysSerThrAlaSerSerThrTrpSer	422
Db	2503	CTAATACAGCAATATATAAGAGCTTCAAAAAGTTAAAGACTCAAGATCTTAACCTCAAT	2562
QY	423	TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val	435
Db	2563	GAAAGCAAGCAAAATACCTTCAAGTCACTACAGCAATTTGACCAAACTTTCCTCGCATTAATT	2622
QY	436	Lys-----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr	453
Db	2623	AAAGCGTATGATCTCTCAAGGAATGCAATTTCTTAAGACAAATAAAGAGCTGAAAAACG	2682
QY	454	ValAsnAsnLeuArg	458
Db	2683	ATTGAGAAATACTATAA	2697
xx	xx	RESULT 11	
xx	xx	ABL29839	
xx	xx	ABL29839 standard; DNA; 3690 BP.	
xx	xx	ABL29839;	
xx	xx	26-MAR-2002 (first entry)	
xx	xx	Drosophila melanogaster genomic polynucleotide seq ID NO 40990.	

RESULT	ID	ABL29839	standard; DNA; 3650 BP.
XX	AC	ABL29839;	
XX	DT	26-MAR-2002 (first entry)	
XX	XX	Drosophila melanogaster genomic polynucleotide	SEQ ID NO 40990

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
OS WO200171042-A2.
PN 27-SEP-2001.
PD 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
PA Venter JC, Adams M, LI PWD, Myers EW;
PI WPI; 2001-656860/75.
DR WPI; 2001-656860/75.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PR interactions -
PS Claim 1; SEQ ID NO 40990; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB872072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3690 BP; 796 A; 1149 C; 949 G; 796 T; 0 other;
Alignment Scores:
Pred. No.: 0.0892 Length: 3690
Score: 138.00 Matches: 107
Percent Similarity: 35.09% Conservative: 73
Best Local Similarity: 20.86% Mismatches: 205
Query Match: 5.73% Indels: 128
DB: Gaps: 21
US-09-825-414-66 (1-487) x ABL29839 (1-3690)
QY 11 ProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 30
DB 496 CCCGAAATCACTAAAGTGGCGACATACACCTCGCCCGGATATAGTTACAGTTCGGCACT 555
QY 31 -----ValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyr 48
DB 556 CCGGAAATCACTAAAGTGGCGACATACACCTCGCCCGGATATAGTTACAGTTCGGCACT 615
QY 49 LeuTyrSHISHISValPheAlaIleAlaHisArgPheSerValIleGlySerProAspGluArg 68
DB 616 GCGCAACCACTGACTTCCAGGTGAGACATACAC-----TCACCT----- 657
QY 69 AspAlaAlaLeuAlaHisAsnGluGlnIleAsp-----AlaLeuValGluThrArgAla 86
DB 658 -----GCGTACACGTACAGCAAGACACACCGGCTATCCAGGTGAGACATACAC 711
QY 87 AsnArgLeuTyrSerGluGlyIuThrProAlaThrIle----- 99
DB 712 TCTCTGTTACAGCTATAGACAGATCTCGCGGCACTCTCCAGATACACCACTACTCG 771
QY 100 -----AlaGluThrPheAlaIleAlaGluIuLysPheAsp----- 110

DB 772 CCATGCGTTTCGTATTCACGCCCAACAGATTGCCAGTGTCCACCTACTACGCGCTTCA 831
QY 111 ---ArgLeuAlaThrThrAlaSer---SerAlaPheGluAsnThrProPheAlaIleAla 128
DB 832 GTGAACCTGGCCACCACTTCGTACTGCTCTCTCCACGGAACAGATACTCCGCAAC 891
QY 129 -----SerValLeuGlnTyrMetGlnProAlaIleAsnIuGlyAspThrLeuAla 145
DB 892 TATGCTCCCTCAATATACAGATACAGCAAGATATAGATATATCCACACAGATATATCC 951
QY 146 ThrProLeuLysProLeuThrProLeuIleSer-----GlyAla 158
DB 952 AAGCTATATGTCGACAGCTATCCGGCATCACCAAAAGTGGCGCCAGCTACGAGGCACT 1011
QY 159 LeuSerGlyAlaMetAspGln-----ValGlyThrLysMetLeuAsp 172
DB 1012 GCATCGGAGCGCTGTCCACCAATATGTTCTCCAAACCAAGCATTCGCAAGTGTCCACT 1071
QY 173 ArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro-----AspLysLeuHisAsp 190
DB 1072 TACGCTGCTCCACGAGTTCACACTTACTATCGGGCTGCCATCGAGCTCTCCAC 1131
QY 191 AlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMet 210
DB 1132 ACCTATGAGCTTCA-----GGATCGGAGCGCTGTCCACCAATATGTTCTCGAAG 1182
QY 211 GlyIleAlaValGlnThrPheSerAlaLeuAsnValArgThrValLeuAlaProAla 230
DB 1188 CCAGCTGTGCTATTTGCCCTCCACAGTGGCAAGGTGGCAAGCTATATGACCTCTGCG 1242
QY 231 LeuAlaSer-----ArgProSerValGlnGlyAlaValAspPheGlyValSerThr 247
DB 1242 ATTCAAGCTACTCCACAGTGTCTCTCCATTGGAAGTAGATAGTCTACGACACTCCACT 1302
QY 248 AlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArg 267
DB 1308 GTCTCCACCTACTCCTCGGATATGTTATGTTGCTG----- 1338
QY 268 AspGlnLeuArgGlyValAlaPheValLeuGlyMetLysAspLysGluProLysAlaAla 287
DB 1339 -----AGTGGTCAAGGGCGCTATCGCACCACTACGTCTCCAGCTCGGCTGGCC 1389
QY 288 LeuSerGluGluThrAspThrLeuAspAlaTyrLysAlaIle---LysSerAlaSerTyr 306
DB 1390 ATTCTCA-----GCCGCTCCAGCATCGCAAGTGGCTACTACTAT 1428
QY 307 SerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu----- 322
DB 1429 GCAGCTCCCGCATTTCACT-----TATGCGGACAGCTCCAGTGGTACCAAAAGTC 1479
QY 322 ----- 322
DB 1480 GCCACTGATATAGAGAAAGTGGCTCGGGATATACATTGGGGCGCTGTTCCCATCATAT 1539
QY 323 -----AspValAlaThrAspGlyLeuLysAlaValArgSerLeu 335
DB 1540 GTCTCTAAGCCAGCGGTGCTAAAGTGGCAACCTATGCTGCTCCGCAATTTCACATATC 1599
QY 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyLysAlaGlyVal 355
DB 1600 TCAGCAGCTCCGGCAGTGCAGATCCAGATCCGACACACTTATGGAGTTCGGGTGATGAGCC 1659
QY 356 SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
DB 1660 GTCTCCCATCATATATATCTCCAGCAGACAGATATGATCTCAGCAGCCCGCA----- 1713
QY 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrAlaGly 395
DB 1714 -----ATCGCTAAGTAGAGACATATAGCTTTCGCGCATATATCAACG 1755
QY 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
DB 1746 TATGCACTGCTCTGTGTGTCAAAAGTGGCTACTACGAGAGCTCTCCATATACCACT 1815

OY 416 ThrAlaSerSerThrThrSerTyValAlaAspGlnThrValLysLeuAlaLysThr--- 434
Db 1816 TACTCTCTCCGCTCCGCT-----CTACCCAAAGGTTTCC 1848
OY 435 -----ValLysAspMetSerGlyAlaLysSer 444
Db 1849 TATAGTCAGCGCAGGATGTATCCACACGATATTTC 1887
RESULT 12
AA67769
ID AAF67769 standard; DNA; 3831 BP.
AC AAF67769;
XX 11-APR-2001 (first entry)
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:53.
XX
XX Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KM membrane construction and membrane transport protein; petroleum spill;
KM hydrocarbon degradation; gram positive aerobic bacterium; marker;
KM identification; microorganism; fine chemical production; transformation;
KM genome mapping; genetic engineering; ds.
XX
XX Corynebacterium glutamicum.
OS
XX WO200100805-A2.
PM
XX 04-JAN-2001.
PD
XX
XX 23-JUN-2000; 2000WO-1B00926.
PF
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
XX (BADI) BASF AG.
PA
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX

DR WPI: 2001-071486/08.
DR P-PSDB: AAB76536.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
XX Claim 3; Page 236-241; 1119pp; English.
PS
XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76847. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.
XX
XX Sequence 3831 BP; 719 A; 1190 C; 1087 G; 835 T; 0 other;
SQ
XX
XX
XX Alignment Scores:
Pred. No.: 0.0935 Length: 3831
Score: 138.00 Matches: 119
Percent Similarity: 37.16% Conserves: 85
Best Local Similarity: 21.68% Mismatches: 219
Query Match: 5.73% Indels: 126
DB: 22 Gaps: 23
US-09-825-414-66 (1-487) x AAF67769 (1-3831)
OY 7 AlaGlnGlnProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 937 GCACACACCGCACCTGACCTGCTGGTGGCCGCGATGCCAGATTTCCTCCGCGTA 996
OY 27 AlaSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaLeuThr 46
Db 997 AAGTCCGCGATCATCCACGACGCGATTCGCGATTCGTTGGTAAGCGATCTGCTCCGTG 1056
OY 47 AspTyrLeuLysHisValPheAlaAlaHisArgPheSerValIleGlySerProAsp 66
Db 1057 GTTCACCTGGACACCACTT-----GTCTCAAGTCCCATGTC 1095
OY 67 GluArgAspAlaAlaLeuAlaHisAsnGlnGlnIleAspAlaLeuValGluThrArgAla 86
Db 1096 CCAACCGACGGCGCGCTGTAAGTTCCGCGACCTGCCAGTAGATGATACCGACCTG 1155
OY 87 AsnArgLeuTyr-----SerGluGlyLeuThrProAlaThr 98
Db 1156 GGTCCGCTCTGTATGTCGCGCGACAGAAATCCGCGACGAGAAAGGCAACCTGATTCCTAGC 1215
OY 99 IleAlaGluThrPheAla---LysAlaGluLysPheAspArgLeuAlaThrThrAlaSer 117
Db 1216 TTGGCTGACGTTTGCATCTCCGCGACGCGAGGCAAGGCAAGCTGTGTCGACGCGCAACAC 1275
OY 118 SerAlaPheGlu-----AsnThrProPheAlaAlaAlaSerValLeuGlnTyrMet 134
Db 1276 TCCGCACTGCCAACAACGCTGACACCCGACCTGACGCTGCGGCAAGCTGTCGACGCTGTT 1335
OY 135 GlnPro-----AlaIleAsnLysGlyAspTyrLeuAlaThrProLeu 148
Db 1336 GCACCTGAATCCATCCGCGCATTCCTCGTATCTCCGCTGAC----- 1377
OY 149 LysProLeuThrProLeu-----IleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 1378 -----CGCAACCCAAATTCACCTGTATGTGGGCTCCCTGTGGTGTGTCGACAGT 1431
OY 165 GlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuGlnIleTyrLeuSerThrSer 184
Db 1432 GTGATGTCGCGACGAGATGTGAGACCTGTGCGATGGTGAAGTATGATGCGCGGTGCACATTC 1491

185 ProasplysleuH1AspAlaMetAlaValSerValLysArgH1S----- 199
1492 AACGATGAGGATCCAAATCCCGCAGCCAAAGTCCGGAAATACCGCAACCATGCTG 1551
200 SerProAlaLeuGlyArgGlnValAlaAspMetGlyLe-----AlaValGlnThr 216
1552 GCACCAAGTTCCTGAGGAAATGCAATTCAGCTTGAGCCCTCCGACAGTGACAAAC 1611
217 PheSerAlaLeuAsnValAlaArgThrValLeuAlaProAlaLeuAlaSerArgProSer 236
1612 CGCCAGGAAATGGAGAGAGTCCGACCCGTTACCGCA-----AAC 1650
237 ValGlnGlyAlaValAlaAspPheGlyValSerThrAlaGlyGlyLeuValAla-----Asn 254
1651 GTCAACGGCAACTTA-----GTGGCTTACCGCACCGCTGTGTGTCGACGTCACATCT 1701
255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAla 274
1702 ACTTTCACGCAATCCAGGCGCAGGCAATTCAGTCCCGCGGC----- 1743
275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluThrAspTrp 294
1744 -----ATGGGTATGAGAGCAGCCGTRACTCTCAGCAGAGCTCGCGCATCTGGAGCCG 1797
295 LeuAspAlaTrpLysAlaAlaLeuSerAlaSerTrpSerGlyAlaAlaLeuAsnAlaGly 314
1798 GCCGATGACACACCCGCAATTAAG-----CTGGGCTCTCCATCTGTG 1839
315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuValAlaValArg--- 333
1840 GAAATTCGTGGAACACCCAGCAAGTAAACCGTGCAGGAGGAGAAATTCCTCCACCCA 1899
334 ---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyLys 352
1900 GACGGCGTTTGTACCTCACCAGTCACTCAGGAGGCGCATGGCAACTCTG---GGCGTT 1956
353 AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnLethrAspSerAlaThrLys 372
1957 GCTCAGATCGCTGAATTCGGGAAGACACATCCCTTGAACCGCGTCACTTGTCTGCGA 2016
373 AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValAlaPheAlaGlyTrpThr 392
2017 CACTCCGTTGGTGAATACAGCGCTTGCCTCATATGCTGTGTGCTCTCCCTGGAAATCC 2076
393 Thr-----AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
2077 GTTCTGAGATCGTTTACCCTGCTGCTTACCATGACACCGCTTGGTGATCGGATGA 2136
407 SerPheLeuGln-----AspLysValLysSerThrAla 417
2137 AACGCTGTGTCACACTACGCGCTCGACACTTCGCCCCAACAGATGGTCTGACCCCA 2196
418 SerSerThrThrSerTrpValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp 437
2197 GACAACGTTTTCGATTACGTTGCG-----TCTGTTTCCGAA 2232
438 MetSerGlyGlu-----AlaLeu 443
2233 GCTTCGGTGAATTCCTCGAGATCGTTAACTACAACCTTGGCTGGCTCAGTACGCAATTT 2292
444 SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaPro 463
2293 GCTGGAAACCCGAGCTGCTT-----GCCGCCCTT 2322
464 GluAlaAspIleGluGlu-----GlyGlyIleSerAlaPhe-----SerArgSer 478
2323 CGTGCAGGATGTTGAGAACCGTGACACAGGTGACGCTTCATTATTTGATCCCTGGCAATT 2382
479 GluThrProPheGlnLeuArgArgLeu 487
2383 GACGTGCCATTCCACTCTCCAAAGCTG 2409
RESULT 13

AA67767
ID AAF67767 standard; DNA: 6136 BP.
AC AAF67767;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:49.
XX
KW Corynebacterium glutamicum; brevbacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN W020010805-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000MO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032190.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041385.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADT) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX WPI: 2001-071486/08.
XX P-PsDB; AAB76534.
XX
XX Corynebacterium glutamicum nucleic acids encoding membrane construction
XX and membrane transport proteins or their portions, useful for typing or
XX identifying C. glutamicum or related bacteria, and as markers for
XX transformation -
XX Claim 3; Page 216-224; 1119pp; English.
XX

CC AAF65774.o: AAF68080 encode the Corynebacterium glutamicum membrane
CC construction and membrane transport (MCT) proteins given in AAB76510 to
CC AAB76647. The MCT nucleic acids and proteins are useful in the
CC identification of microorganisms which can be used to produce fine
CC chemicals, for modulating fine chemical production in C. glutamicum or
CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
CC identification of C. glutamicum or related bacteria, as reference points
CC for mapping C. glutamicum genome, and as markers for transformation.
CC AAF68082 and AAF68082 represent sequencing primers which are used in an
CC example from the present invention.

Sequence 6136 BP; 1182 A; 1840 C; 1754 G; 1360 T; 0 other;

Alignment Scores:

Pred. No.:	0.17	Length:	6136
Score:	138.00	Matches:	119
Percent Similarity:	3.15%	Conservative:	85
Best Local Similarity:	21.66%	Mismatches:	219
Query Match:	5.73%	Indels:	126
DB:	22	Gaps:	23

US-09-825-414-66 (1-487) x AAF67767 (1-6136)

OY	7	AleGlnElmPrProGlValAlaMetGIUserPheArgThrAlaSerAspAlaSerIeu	26
Dd	3083	GCAACACCAGCCACTACAGCTGGTGTGGCCCCCGCATGGCCACAGTTTCCTGCCCTTA	31422
OY	27	AlaserSerSerValArgSerValSerThrThrSercysArgAspLeuGlnAlaIlethr	46
Dd	3143	AAATGCAGCAGTACATCCAGGACCCAGATTCCGCATCCGTTGTGAAGCATCTGTCCCTG	3202
OY	47	AsPTyrLeuLysHisIstValPheAlaAlaHisArgPheSerValIleGlySerProasp	66
Dd	3203	GTTCACCTGGAGGACACACAATT-----GTCTCAATGTCGATGTC	3241
OY	67	GLuArgSprAlaIleValAlaHisAsnGluGlnIleAspAlaLeuValGlnThrArgAla	86
Dd	3242	CCAACGACGAGCGCGCTGGAAGCTTTCCGCGACTGCCGAGAGTGATGATACGACCTG	3301
OY	87	AsnArgLeuLyr-----SerGlnGlyIstThrProAlathr	98
Dd	3302	GGTGGCTGTGATCTGGGGCGGAGAATAAGGCGGAGGACGACCACTATGCTTAGC	3361
OY	99	IleAlaGluThrPheAla---LysAlaGluLysPheAspArgLeuAlaThrThrAlaser	117
Dd	3362	TGTGGCTGAAGGTGTGCGATGCGCGGACGSCGACGAGGACGCGTGTGCGCACCAACAC	3421
OY	118	SerAlaPheGlu-----AsnThrProPheAlaAlaAlaSerValLeuGlnIutMet	134
Dd	3422	TTCGCACTGGCAACACACGCTGGACACSSCACTGACTGCGCGCATGGACACGCTGTT	3481
OY	135	GlnPro-----AlaIleAsnLysGluAspTrpLeuAlaThrProIeu	148
Dd	3482	GCACTGATATCSATGCGGCCATCGCTGATGATCCGGTAC-----	3523
OY	149	LysTrpLeuThrProIeu-----IleSerGlnAlaLeuSerGlnAlaMetasp	164
Dd	3524	-----CGAACCCAAATACSTGTCTGTATGTGGGGTCCCTGCGTGTGCGCAGT	3577
OY	165	GlnValGluThrLysMetMetAspArgAlaArgGlnLysPheIstLysLeuSerThrSer	184
Dd	3578	GTATGCTGCGAAGCGATGTGGACSTGTGSCATTCGGTAAGTAATGATGGCCGGTGCACATC	3637
OY	185	ProAspLysLeuHisAspAlaMetAlaValSerValLysValGnis-----	199
Dd	3638	AACATATGACGAGATCCAAATCCSSCGAGSCAAGCTGTGGATATACACSSCAACATCTGTG	3697
OY	200	SerTrpAlaLeuGluArgGlnValAlaAspMetGlyIle-----AlaValGlnThr	216
Dd	3698	GCAACAGTTTCTTCACAGSTGAAGAAATTGGATTCCAGCGTGAAGCCTCCSGCATGGACAC	3757
OY	217	PheSerAlaLeuAsnValAlaArgThrValLeuAlaProAlaLeuAlaSerArgProser	236

[illegible]

KW pharmaceutical; gene; ds.
XX Drosophila melanogaster.
XX WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
XX 11-JUL-2000; 2000US-0614150.
XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PMD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions -
XX
XX Claim 1: SEQ ID NO 40987; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signaling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB101840-AB16175) and the encoded proteins
XX (AB57737-AB872072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6169 BP; 1607 A; 1369 C; 1664 G; 1529 T; 0 other:
SQ

Alignment Scores:
Pred. No.: 0.172 Length: 6169
Score: 138.00 Matches: 107
Percent Similarity: 35.09% Conservative: 73
Best Local Similarity: 20.86% Mismatches: 205
Query Match: 5.73% Indels: 128
DB: 23 Gaps: 21
US-09-825-414-66 (1-487) x ABL29838 (1-6169)
QY 11 ProGlyValAlaIaMetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 30
DB 4325 CCCGGAATCACTAAAGTGGACATACCTCGCCGATATGTTACAGTTGGCAAGT 4266
QY 31 -----ValArgSerValSerThrThrSerCysArgAspLeuAlaIleThrAspTyr 48
DB 4265 CCGGGAATTCACAAAGTGGACATACCTCGCCGATATGTTACAGTTGGCAAGT 4206
QY 49 LeuLysHisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArg 68
DB 4205 GCGCCACCGATCTCCAAAGTGGACATACAGC-----TCACCT----- 4164
QY 69 AspAlaAlaLeuAlaHisAsnGluGlnIleAsp-----AlaLeuValGluThrArgAla 86
DB 4163 -----GCGTACACGTCACAGCAAGACACACCCGGCTACTCCAAAGTGGACATACAGC 4110
QY 87 AsnArgLeuThrSerGluGluGluThrProAlaThrIle----- 99
DB 4109 TCTCCGTATACAGTATGACAGATCTCGCCGGCATCTCCAGATAGCAGCTACTCG 4050
QY 100 -----AlaGluThrPheAlaAlaGluAlaGluLysPheAsp----- 110
DB 4049 CCATCGTTTCGTATTCACGCGCAACGATGCCAAGTGTCCACCTACTCAGCGCTTCA 3990

QY 111 ---ArgLeuAlaThrThrAlaSer---SerAlaPheGluAsnThrProPheAlaAla 128
DB 3989 GTGAAMAATGGCCACCACTTGTCTCTCTCTCCACGAAACAGATATCCCGCAC 3930
QY 129 -----SerValLeuGlnIleThrMetGlnProAlaIleAsnLysGlyAspThrLeuAla 145
DB 3929 TATGTCCTCCATATACCAAGTACACAGCATGATGATATCCACCAAGTATATCC 3870
QY 146 ThrProLeuLysProLeuThrProLeuIleSer-----GlyAla 158
DB 3869 AAGCCTATATGAGCAGACCTATCCGGGATCACAAGTGGCCGACATCAGGAGCACT 3810
QY 159 LeuSerGlyAlaMetAspGln-----ValGlyThrLysMetAsp 172
DB 3809 GCATCCGAGCGGTGTCACACAGTATGTCACCAACGACATTCGCAAGTGTCCACT 3750
QY 173 ArgAlaArgGlyAspLeuHisThrLeuSerThrPro-----AspLysLeuHisAsp 190
DB 3749 TACGCTGCTCCACGAGCTCAACTATCTATGCGCTGCGCATGTCGAAGCTCCACC 3690
QY 191 AlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMet 210
DB 3689 AGCTATGAGACTTCA-----GGATCGGAGCCGCTCCACCAATATGTCTCGAAG 3639
QY 211 GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAla 230
DB 3638 CCAGCTGTGCTATTCCTCCCTCCACGATGGCCAGATGGCAAGTATGACCTCTCTG 3579
QY 231 LeuAlaSer-----ArgProSerValGlnGlyAlaValAlaAspPheGlyValSerThr 247
DB 3578 ATTTCAGCTACTCCACTGCTGCTGCTGCAATTCGAAGTATGATCTACGACCTCCACT 3519
QY 248 AlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArg 267
DB 3518 GTCTCCACTACTCCCTCCGGATATGTTATGTTGCTG----- 3483
QY 268 AspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAla 287
DB 3482 -----AGTGGCTCAGAGGCGCCGATCGCACAGTACGTGTCCAGCTCGGTGCGC 3432
QY 288 LeuSerGluGlnThrAspThrPheLeuAspAlaTyrLysAlaIle---LysSerLysSerTyr 306
DB 3431 ATTTC-----GCCGCTCCAGCATGCGCAAGTGGCTACTCTAT 3393
QY 307 SerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu----- 322
DB 3392 GCAGCTCCCGCATTTCACT-----TATGCGCAGCTCCAGTGTGACCAAGTC 3342
QY 322 ----- 322
DB 3341 GCCACTGATATGAGGAAGTGGCTCGGATACAGTTCCGGCGCTGTTCCCATCAATAT 3282
QY 323 -----AspValAlaThrAspGlyLeuLysAlaValAlaArgSerLeu 335
DB 3281 GTCTCTAAGCCAGCGGTGCTAAAGTGGCAACCTATGCTGCTGCTCCATTCACATATC 3222
QY 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyLysAlaGlyVal 355
DB 3221 TCAGCAGCTCCGCGACGTGACCAAGATCGGACAAAGTATGAGAGATCGGATGAGGCC 3162
QY 356 SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
DB 3161 GTCTCCCATCAGTATGTCTCCAAAGCAGAGATGAGCATTCAGACGCCCGCCCA----- 3108
QY 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGly 395
DB 3107 -----ATCGCTAAGGTAGGACATACAGCTTCTCCGCAATATACACAG 3066
QY 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
DB 3065 TATGCCACTGCTCTGCTGCTCTCAAAAGTGGCTATCAGCAGCTCCCTCCATFAGCACT 3006
QY 416 ThrAlaSerSerThrThrSerThrValAlaAspGlnThrValLysLeuAlaLysThr--- 434

Db 3005 TACTCTCGCTCCGCGCT-----CTAGCCAAAGTTTCC 2973
OY 435 -----ValLysAspMetSerGlyAlaIleSer 444
Db 2972 TATAGTCAGCAGCGAGTATCCACCACTATTC 2934
RESULT 15
AAH67704
ID AAH67704 standard; DNA; 8979 BP.
XX
AC AAH67704:
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 2739.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
FN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PE 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOSYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
DR P-PSDB: AAG92485.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 8; SEQ ID NO: 2739; 246bp + Sequence Listing: English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX
SQ Sequence 8979 BP; 1749 A; 2763 C; 2544 G; 1923 T; 0 other;
Alignment Scores:
Pred. No.: 0.277 Length: 8979
Score: 138.00 Matches: 119
Percent Similarity: 37.16% Conservative: 85
Best Local Similarity: 21.68% Mismatches: 219
Query Match: 5.73% Indels: 126
DB: 22 Gaps: 23
US-09-825-414-66 (1-487) x AAH67704 (1-8979)

OY 7 ALaGInGluProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 3073 GCACACACCGCAGCTGACGTGCTGGCCCGCATGCCACAGCATTTTCCTGCCGTA 3132
OY 27 AlaSerSerValArgSerValSerThrThrSerCysArgAspGluAlaIleThr 46
Db 3133 AGTCCGCGATGATCCAGGACCGCATTCGGATCGCTTGTAAGGACATGCTGCCG 3192
OY 47 AspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAsp 66
Db 3193 GTTCACCTGGACACCACTT-----GTGCTCAAGTCCCATGTC 3231
OY 67 GluArgAspAlaAlaLeuAlaHisAsnGluInLeuAspAlaLeuValGluThrArgAla 86
Db 3232 CCACCGACGGCGCGGTGAAGTTTCGGGACTGCCGATGAGTGTGATCCAGCTG 3291
OY 87 AsnArgLeuTyr-----SerGluGlyLeuThrProAlaThr 98
Db 3292 GGTGCGCTGTGATGTGGCGGCAGAAATCGCCGACGAGAGGACCACTGATGCG 3351
OY 99 IleAlaGluThrPheAla----LysAlaGluLysPheAspArgLeuAlaThrAlaSer 117
Db 3352 TTGGCTGAGCGTTTCCGATCCGCGAGCGCAAGGCGCAAGCGTGTGCGACCAACAC 3411
OY 118 SerAlaPheGlu-----AspThrProPheAlaAlaSerValLeuGluInTyrMet 134
Db 3412 TCCGCACTGCGCAACCAAGCGTGGACATGCTGATGATCGCGGTAAC----- 3471
OY 135 GluPro-----AlaIleAsnLysGlyAspTyrLeuAlaThrProLeu 148
Db 3472 GCACCTGAATCCATGCGGCGCATGCTGATGATCGCGGTAAC----- 3513
OY 149 LysProLeuThrProLeu-----IleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 3514 -----CGCAACCCAAATGATCAGCTGTGATGATGCGGCTGCTGCTGCTGCGAGCT 3567
OY 165 GluValGluThrLysMetMetAspArgAlaArgGluAspLeuHisTyrLeuSerThrSer 184
Db 3568 GTGATCGTGCACGCGATGTGGACCTGTGCAATGCGTGAATGATGCGCGGTGCGACATTC 3627
OY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis----- 199
Db 3628 AACGATGAGCAATCCAAATCCCGCAGCAGAGTGTGATGATACCGCAACCATGCTG 3687
OY 200 SerProAlaLeuGlyArgGlnValValAspMetGlyIle-----AlaValGlnThr 216
Db 3688 GCACCAAGTCTTCCAGTGAAGAAATGATGATCAGCGTGGACGCTGCCAGTGAACAC 3747
OY 217 PheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSer 236
Db 3748 CCGCCAGGAATGGAGAGGTGCGCACCTTACCGCA-----ACC 3786
OY 237 ValGlnGluValAlaAspPheGlyValSerThrAlaGlyLeuValAla-----Asn 254
Db 3787 GTCAACGCGCACTT-----GTGCTTACCGCCACCGCTGTGTGCGACCTCCATCT 3837
OY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAla 274
Db 3838 ACTTCTACGCAATCCACAGGCGCAGGCAATTCAGTCACTGCTCAAGGC----- 3879
OY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluInThrAspThr 294
Db 3880 -----ATGGGTATGGAAGCAGCGCGTAACTCAAGCAGCTCGCGGTATGCGGACCGC 3933
OY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 3934 GCCGATGACACACCGCAATGAG-----CTGGGCTTCTCCATCGTG 3975
OY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArg--- 333
Db 3976 GAATTCGTGAAACACCAACCGCAGAGTACCGTGGCAGGAGAGTCTTCCACACCA 4035
OY 334 ---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyLys 352

```
Db 4036 GACGCGCTTTTGTACTACCTCACCACCTTCCACCAGGTGGGCATGGCAACTCTG---GGCGTT 4092
      ::::: ||| ||::: ||::: |||
QY 353 AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnLleThrAspSerAlaThrLys 372
      ||| ::::: |||
Db 4093 GCTCAGATCCGTGAATGCGTGAAGACACATCCCTTGAACACCGCGCATCTTGTGCTGA 4152
      ::::: |||
QY 373 AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThr 392
      ::::: |||
Db 4153 CACCCCGTGTGAGTGAACAAGCGCTGCTCATATGCTGTGCTGTCCCTGGAATCC 4212
      ::::: |||
QY 393 Thr-----AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
      ::::: |||
Db 4213 GTTCTGAGATCGTTTACCGTCTGCGCTGACATGCACCGCTTGGTGTGATCGCGATGAA 4272
      ::::: |||
QY 407 SerPheIleGln-----AspLysValLysSerThrAla 417
      ::::: |||
Db 4273 AACGGTCTGTCCACTACAGCGCGCTGCAGCCTCTGCCCCCAACAAAGATGGTCTGACCGCA 4332
      ::::: |||
QY 418 SerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp 437
      ::::: |||
Db 4333 GACACACGTTTCGATTCAGTTGCG-----TCTGTTCCGAA 4368
      ::::: |||
QY 438 MetSerGlyGlu-----AlaIle 443
      ::::: |||
Db 4369 GCTTCCGCGTAATTCCTGAGATCGTTTAACTACAACCTTGCGCTGCCTGACGACGAGTT 4428
      ::::: |||
QY 444 SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaPro 463
      ::::: |||
Db 4429 GCTGGAACCCAGCGCTGCTT-----GCCGCCCTT 4458
      ::::: |||
QY 464 GluAlaAspIleGluGlu-----GlyGlyLleSerAlaPhe-----SerArgSer 478
      ::::: |||
Db 4459 CGTCCCGATGTTGAGAACCGTGCACACGATCAGCGTGCCTTCATTTGATCCCTGCATTT 4518
      ::::: |||
QY 479 GluThrProPheGlnLeuArgLeu 487
      ::::: |||
Db 4519 GACGTGCCATTCACATCCCTCCACAGCTG 4545
```

Search completed: January 31, 2003, 05:30:03
Job time : 293.246 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 31, 2003, 03:21:14 ; Search time 1575.62 Seconds

(without alignments)
5005.784 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407

Sequence: 1 MHINQSAQPGVAMESFRT.....EEGISAFSRSEPFQRLRL 487

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+P2n.model DEV=alp
-Q/cgn2_1/USPFO.spool/US09825414/runat_28012003_161742_26093/app_query.fasta_1.1294
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-USER=US09825414.cgn_1_1_2221_@runat_28012003_161742_26093 -NCPV=6 -ICPV=3
-NO_XLPXY -NO_MMAB -LARGEQUERTY -NES_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_liv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136	5.7	2191	AQ012154	AQ012154 8758PID03
C 2	128	5.3	883	BG369810	BG369810 HVSME1002
C 3	122	5.1	4001	AK014542	AK014542 Mus muscu
C 4	115.5	4.8	700	AL508808	AL508808 AL508808
C 5	113.5	4.7	1094	CNS076CM	AL433228 T3 end of
C 6	111.5	4.6	833	BQ800257	BQ800257 EST 7292
C 7	107	4.4	680	AL508002	AL508002 AL508002
C 8	107	4.4	842	BQ800337	BQ800337 EST 7372
C 9	107	4.4	1150	CNS0792S	AL433950 T3 end of
C 10	107	4.4	1763	AY107053	AY107053 Zee mays
C 11	105.5	4.4	841	BQ800379	BQ800379 EST 7414
C 12	105.5	4.4	850	BF265543	BF265543 HV_CBE001
C 13	105.5	4.4	888	CNS076CP	AL431231 T7 end of
C 14	105	4.4	579	BQ795937	BQ795937 EST 4875
C 15	105	4.4	806	BM077210	BM077210 TREST-A41
C 16	105	4.4	853	BQ800263	BQ800263 EST 7298
C 17	105	4.4	886	BQ794648	BQ794648 EST 3586
C 18	104.5	4.3	842	BQ792279	BQ792279 EST 8734
C 19	104.5	4.3	965	BQ674550	BQ674550 AGENCOURT
C 20	104	4.3	587	BE583087	BE583087 3-5A-MY P
C 21	104	4.3	858	BF265524	BF265524 HV_CBE001
C 22	103.5	4.3	613	AT517569	AT517569 CH2575.5
C 23	103.5	4.3	842	BQ797946	BQ797946 EST 6884
C 24	103.5	4.3	2779	AK004928	AK004928 Mus muscu
C 25	103	4.3	718	BQ799997	BQ799997 EST 2166
C 26	103	4.3	743	BQ793911	BQ793911 EST 2849
C 27	103	4.3	774	B1556804	B1556804 603238734
C 28	103	4.3	778	BQ799265	BQ799265 EST 1434
C 29	103	4.3	800	BQ794248	BQ794248 EST 3186
C 30	103	4.3	821	BQ796730	BQ796730 EST 3568
C 31	103	4.3	871	BQ794592	BQ794592 EST 3545
C 32	103	4.3	933	BF065640	BF065640 HV_CBE001
C 33	103	4.3	1034	CNS06060	AL410284 T3 end of
C 34	103	4.3	1068	BF540337	BF540337 602052394
C 35	102.5	4.3	2539	AK004578	AK004578 Mus muscu
C 36	102	4.2	1009	BM904265	BM904265 AGENCOURT
C 37	101.5	4.2	681	BM582171	BM582171 17006872
C 38	101.5	4.2	989	BQ065543	BQ065543 AGENCOURT
C 39	101.5	4.2	1031	CNS06042	AL393369 T3 end of
C 40	101.5	4.2	1138	CNS060DE	AL410542 T3 end of
C 41	101	4.2	570	AM553832	AM553832 L0232D01-
C 42	101	4.2	691	AM553048	AM553048 L0221B12-
C 43	101	4.2	763	AQ012153	AQ012153 87581A12
C 44	100.5	4.2	642	B1952036	B1952036 HVSME000
C 45	100.5	4.2	649	B1958467	B1958467 HVSME001

ALIGNMENTS

RESULT 1
LOCUS AQ012154 2191 bp DNA linear GSS 04-JUN-1998
DEFINITION 8758PID037011098 Cosmid library of chromosome II Rhodobacter
ACCESSION AQ012154
VERSION AQ012154
KEYWORDS GI:3177109
SOURCE GSS.
ORGANISM Rhodobacter sphaeroides
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
Rhodobacteraceae; Rhodobacter.
REFERENCE 1 (bases 1 to 2191)
Choudhary,M., Mackenzie,C., Mouncey,N., Weinstock,G.M. and Kaplan
,S.

TITLE	RSGBD, the Rhodobacter sphaeroides Genome Database
JOURNAL	Unpublished (1998)
COMMENT	Contact: Choudhary, M.

Department of Microbiology and Molecular Genetics
University of Texas Medical School
6431 Fannin Street, Houston, TX 77030, USA
Tel: 713 500 5437
Fax: 713 500 5499

Email: madhu@utmmg.med.uth.tmc.edu
Seq primer: pBluescript T7
Class: shotgun.

FEATURES	Location/Qualifiers
source	1. . 2191

```

/organism="Rhodospirillum rubrum"
/strain="2.4.1p"
/db_xref="taxon:1063"
/clone="8758p1D037011098"
/clone_lib="Cosmid library of chromosome II"
/lab_host="E. coli S17-1"
/note="Vector: pUA2917"
BASE COUNT      341 a      749 c      718 g      382 t      1 others
ORIGIN

```

Alignment Scores:	
Pred. No.:	0.00649
Score:	136.00
Percent Similarity:	37.58%
Best Local Similarity:	25.28%
Query Match:	5.65%
DB:	17
Length:	2191
Matches:	113
Conservative:	55
Mismatches:	163
Indels:	117
Gaps:	21

US-09-825-414-66 (1-487) x AQ012154 (1-2191)

OY	95	ThrProAlaThrIleAlaGlnThrProAlaLysAlaGlnLysPheAspArgLeuAlaThr	114
Db	1781	ACGACACATCAACCCCTGGACAGCGAAATTC-----GCCGCCGAATGGAC-----GCCCAT	1734
OY	115	ThrAlaSerSerAlaPheGlnAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMet	134
Db	1733	GTGGCGGCGCTGGGGGCGGCGGAAACCCCTGGAGGCG-----ATC	1695
OY	135	GlnProAlaIleAsnLysGlnAspTyrPheAlaThrProLeuLysProLeuThrProLeu	154
Db	1694	CGCCATCTGGTGGCGCCCGGG-----CTGGCGCGCGGGCCCGCAAACTGCCACACGCTC	1638
OY	155	IleSerGlnAlaLeuSerGlnAlaMetAspGlnValGlnThrLysMetMetAspArgAla	174
Db	1637	TGCATCGGGCTGATCTCTCCGCGATGGAGACAGACCCCTCCGGAAGCTTTCGGCCGGGTT	1578
OY	175	ArgGlnAspLeuIleArgTyrLeuSerThrSerProAspLysLeuIleAspAlaMetAlaVal	194
Db	1577	TGGCGGAGGCTGGGGGTCTGGCCCGGACAGACAGATCGCTTAATACAGAGCTCCCT	1518
OY	195	SerValLysArg-----HisSerProAlaLeuGlnArgGlnValAsp	209
Db	1517	CAATACCCGGGAGACGATCGACTCGCTCATGCTCT-----	1482
OY	210	MetGlyIleAlaValGlnThrPheSerAlaAlaAsnValValArgThrValLeuAlaPro	229
Db	1481	-----GCAAGTCGAACAACAGAGGTGGTCTCAAGCTGC	1449
OY	230	AlaLeuAlaSerArgPro-----SerValGlnGlnAlaValaAspPheGlnValaSerThr	247
Db	1448	GCACCGTCGTGGAGGCGGTGAACCGGCTGTCTGGGGCCCTACG-----	1404
OY	248	AlaGlnLysLeuValAlaAsnAlaGlnPheGlnAspArgMetLeuSerValGlnSerArg	267
Db	1403	-----GGCCTGAGCGCCCGGTCTGGCGCCACAGTCGAGACAGATGTCGGTGAATCCATG	1350
OY	268	AspGlnLeu-----ArgGlnGlnAlaPheValLeuGly-----Met	279
Db	1349	GAGCAAGATTTCGGCCACGGGCGGCGGCTCTCCGCGGTACCGGATCCAGAGCAAGCTC	1290

[illegible]

COMMENT	On Mar 8, 2001 this sequence version replaced gi:13258911.
---------	--


```

Db 1908 CCGGGCAGCTGGCAGAAAAGAGTCTCAGAGCAGGGGCTTCGGCAGAGACTGTGTAT 1967
QY 110 AsparGLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSer 129
Db 1968 GAGCAGATGTGGCGTGTGGAGT-----CGAGCGCCGAG 2003
QY 130 ValLeuGlnTyrMetGlnProAlaIleAsnLys---GlyAspThrLeuAlaThrProLeu 148
Db 2004 GCAGAGCCATCTCTACAGATGCAGTGCAGCAGCAGCTGAGACCC-CTGCACCTCCGCTG 2062
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeuSerGly-AlaMetAspGlnValGlyThr 168
Db 2063 CACCACTCTCCAGACTACTGTGTGAGCCGGCTCAGCAGCCCTGAGACAGCTGAGC-- 2120
QY 168 rlysmetmetasparGAlaArgGlyAspLeuHisTyrIleSerThrSerProAspLysLe 188
Db 2121 -----GGCCTGGAGAGGGCCACACCCAGCTGGCTCTCTCTGAAATGCT-- 2168
QY 188 uhlAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlnYArgGlnValVa 208
Db 2169 ----TCTGCCCTGTGGCAGCGCTGACCGGCTTCTCCATTGGCTGCGGACACCTTGT 2224
QY 208 lAspmetGlyIleAlaValGlnThrPheSer----- 218
Db 2225 CAAT---GGTGGCGCCACCTCCACCTGGCCCGCCAGCAGCCCGCCTGTATGA 2281
QY 219 -----AlaLeuAsnValVal-----ArgTh 225
Db 2282 CACATGCAGGAGGTGGAGCCCGGCTCTGGAGCTGTGGAGCAGCTGCAGACACAGAC 2341
QY 225 rValLeuAlaProAlaLeuAlaSer-----ArgProSerValGlnGlyAlaValAspPh 243
Db 2342 AGTGCTACGAGGGCTCAGCCAGCCTGTATGGCGGGCCCGCTCAGAGGCTTGTGAGATT 2401
QY 243 eglValAspThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSe 263
Db 2402 GGGCCAGGACTTG------AAGCCTAAGAG 2425
QY 263 rValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysG 283
Db 2426 CCGGATGTACCGCAGAGAGAGAGTACGG-----GCCATGTGTGACAGAG 2470
QY 283 uProLysAlaAlaLeuSerGlnGlnThrAspThrPheAspAlaTyrLysAlaIleLysSe 303
Db 2471 G---ATGGCGGCACCTCGCAGACCATGTAG-----GACGCTGTGCGGAGATCGAGGA 2521
QY 303 rAla-----SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgHe 317
Db 2522 CATGATGAGCCAGGCCCGCCAGAGAGAGCTCGGCGTGAACCTGAGCTGATGAGAGAT 2581
QY 317 tAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuVal-- 336
Db 2582 C-----CTCACTCTCCGCACAGACCTGATGAAAGGCTATCCGGCTCTGTGTAT 2629
QY 337 -SerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyLysAlaValSe 356
Db 2630 GACCTCAGCAGCTGCAGAGAGAA-----ATTGTGAGAGAGGCGAGGGGGGCAC 2680
QY 356 rLysLeuGlnLysMetAlaThrLysAsnIle-----ThrAsp-----SerAlaTh 371
Db 2681 AACGCGACGGAATTTTATTCACAGAAATTCACGGTGTGACTGAAGGCTCATCTCACCCTC 2740
QY 371 rLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTr 391
Db 2741 TAAAGCA-----GTGGGCTG 2755
QY 391 pThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlySerPheIleGlnAs 411
Db 2756 GGGAGCCACAGACTG-----GTGGAGTCAAGCTGACAGAGCTTGTGCTTCA 2800
QY 411 pLysValLys-----SerThrAlaSerSerThrThrSe 422
Db 2801 CATGGCAATATGAGAACTATCTGTCTTCATGATGAGATTGCGGCGAGCAGCGGCCCA 2860

```

```

QY 422 rTyrValAlaAspGlnThrValLys----- 430
Db 2861 GCTGTGGCAGCTCTGAGAGTGAAGCCCAAGAAAGACTCCCGCAGCTGAGCCGCTGCA 2920
QY 431 ----LeuAlaLysThrValLysAspMetSerGlyAlaAlaIleSerThr 446
Db 2921 GGAATCTCCGACACTGTCAAGAGAGGGCTGCCAAGCTGTGCTCCAC 2972

RESULT 4
AL508808/c 700 bp mRNA linear EST 04-JAN-2001
LOCUS AL508808 Hordeum vulgare Barke developing caryopsis (3..-15.DAP)
DEFINITION Hordeum vulgare cDNA HY09N21V 5', mRNA sequence.
ACCESSION AL508808
VERSION AL508808.1 GI:12035311
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 700)
REFERENCE
AUTHORS Michael, M., Weschke, M., Pleisner, K., P. and Graner, A.
TITLE EST sequencing and analysis in barley
JOURNAL Unpublished (2000)
COMMENT
Contact: Michael M
Institute for Plant Genetics and Crop Plant Research
Corrensstr. 3, D-06466 Gatersleben, Germany
Email: michael@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de
Seq primer: r3 primer for 5' end.

FEATURES
source
1..700
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HY09N21V"
/clone_1lb="Hordeum vulgare Barke developing caryopsis (3..-15.DAP)"
/clone_15b="Hordeum vulgare Barke developing caryopsis (3..-15.DAP)"
/issue_type="developing caryopsis (3..-15.DAP)"
/lab_host="XL0LR"
/notes="Vector: Plasmid pRK-CMV; Site_1: EcoRI; Site_2: XhoI; mRNA was made from developing caryopsis (3..-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'- and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp."

BASE COUNT 152 a 204 c 244 g 93 t 7 others
ORIGIN

Alignment Scores:
Pred. No.: 0.0195 Length: 700
Score: 115.50 Matches: 45
Percent Similarity: 43.75% Conservative: 25
Best Local Similarity: 28.12% Mismatches: 61
Query Match: 4.80% Indels: 29
DB: 9 Gaps: 6

US-09-825-414-66 (1-487) x AL508808 (1-700)

QY 311 LeuAsnAlaGlyLysArgMetAlaGly-----LeuProLeu 322
Db 664 GTTCCCTTGGGCTCGGCTCGGGCCGAGACACTTCTTTCTCTGCTCTTG 605
QY 323 AspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThr 342
Db 604 -----GTCGCGGCGCTCAAGCGGCTG-----CTTGCCTCAACNGCTNCAC 563

```

QY 343 LysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAla 362
 Db 562 AACTTCGGTCTCTGT---TCCCTCGGCGAGGGGACAGCTCCGGCTTTGGTCTCGGTCA 506
 QY 363 ThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuAla 382
 Db 505 ACTGTCGGCGTGACAGGGGCGCTCCATCGCGGCGGCCGTGACACAACTCTGTCTCG 446
 QY 383 GlySerVal-----GlyValPheAlaGlyTyrThrThrAlaGly 395
 Db 445 GCCTCTGTGGTCCGCGTGGAGCTGCNTCGGCTCTTTCGACGCTCGGTCAAGCGCGC 386
 QY 366 LeuAlaThrAspProAlaValLysLysAlaGlySerPheIleGlnAspLysValLysSer 415
 Db 385 GCAGCAACT-----TCCCTCTTCAGCGCGCTCCGCGCAGGACA 347
 QY 416 ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrVal 435
 Db 346 ACGCGCGGCTCGACAGCTCGACCGCTTGCGACCTCGCTGTCGACGCGCGCTCGACA 287
 QY 436 LysAspMetSerGlyGlyAlaIleSerSerThrGlyAlaSerLeuAlaGlySerThrValAsn 455
 Db 286 ATCTCCACTGCTGCCACCGCTAGCTCTCTCGACCGGACNAGCGCGCTCGACGACTCC 227

RESULT 5
 CNS076CM/C 1094 bp DNA linear GSS 07-JUL-2001
 LOCUS T3 end of clone BB0AA002F11 of library BB0AA from strain CBS 4732
 DEFINITION of *Pichia angusta*, genomic survey sequence.
 ACCESSION AL431228
 VERSION AL431228.1 GI:12214640
 KEYWORDS GSS.
 SOURCE *Pichia angusta*.
 ORGANISM *Pichia angusta*.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; Saccharomycetaceae; *Pichia*.
 1 (bases 1 to 1094)
 Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
 Bortolin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
 deMontigny,J., Dujon,B., Durrens,P., Leplingle,A., Llorente,B.,
 Malpertuy,A., Neuvéglise,C., Ozier-Kaloponitis,O., Potier,S.,
 Sarrin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
 Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of
 yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 JOURNAL MEDLINE 20584711
 PUBMED 11152876

REFERENCE 2 (bases 1 to 1094)
 Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
 and Dujon,B.
 Genomic exploration of the hemiascomycetous yeasts: 13. *Pichia*
angusta
 FEBS Lett. 487 (1), 76-81 (2000)
 JOURNAL MEDLINE 20584723
 PUBMED 11152888
 TITLE 3 (bases 1 to 1094)
 Genoscope.
 Direct Submission
 Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 This GSS is part of a random genomic sequencing program of thirteen
 yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces*
exiguus, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,
Saccharomyces kluyveri, *Kluyveromyces thermotolerans*, *Kluyveromyces*
lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*
angusta, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,
Candida tropicalis and *Yarrowia lipolytica*. Genomic inserts of 3 to
 5 kb were prepared and both extremities were sequenced. See
 keywords for description of this sequence and for the sequence of
 the other extremity of this insert.

FEATURES

Location/Qualifiers

source 1. 1094
 /organism="Pichia angusta"
 /strain="CBS 4732"
 /db_xref="taxon:4905"
 /clone="BB0AA002F11"
 /clone_11b="BB0AA"
 /note="end : T3"
 BASE COUNT 320 a 245 c 321 g 207 t 1 others
 ORIGIN
 Alignment Scores: 0.0719 Length: 1094
 Pred. No.: 113.50 Matches: 94
 Score: 33.05% Conservative: 63
 Percent Similarity: 19.79% Mismatches: 163
 Best Local Similarity: 4.72% Indels: 155
 Query Match: 17 Gaps: 17
 US-09-825-414-66 (1-487) x CNS076CM (1-1094)
 QY 5 GlnSerAlaGlnGlnProGlyValAlaMetGlnSerPheArgThrAlaSerAla 24
 Db 1035 CAGACTTCTAGCCAGACACCTCTCCAAAGCCAACTCA-----ACTTCCACAGCGCT 982
 QY 25 SerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAla 44
 Db 981 TCAGGTGATTCAGCTCTACGTCTGCTGCTTCA---AGTAATGCTCTGACTGTGATACC 925
 QY 45 IleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySer 64
 Db 924 ATTACC-----AGATCTGGCTCAAGTACGATCGGCGCTCG 889
 QY 65 ProAspGlnArgAspAlaAlaLeuAlaHisAsnGlnGlnIleAspAlaLeuValGlnThr 84
 Db 888 TCCACCGGTTCTAGCTCTGCGCACAGTTATGCA----- 856
 QY 85 ArgAlaAsnArgLeuTyrSerGlnGlyGlnThrProAlaThrIleAlaGlnThrPheAla 104
 Db 855 -----TCTAGACATGCTATCTCGATTCGATCCGGAACCTTCCAGC 820
 QY 105 LysAlaGlnLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGlnAsnThrPro 124
 Db 819 TCTGCT-----ACCTCAACTCTCAGGTTAT-----ACGTCA 790
 QY 125 PheAlaAlaAlaSerValLeuGlnIleTyrMetGlnProAlaIleAsnLysGlyAspThrLeu 144
 Db 789 AGCGCGGCACCTCTGATTTAGTTATGGGTACTTCAAGCGGTCAAGTTCCAGCTCT 730
 QY 145 AlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164
 Db 729 TCCAGCTCTGCCACTCTCGATCAAGCCTTTTCACTCTCTGCTCGGCTCCGCTGCC 670
 QY 165 GlnValGlyThrLysMetMetAlaPargAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
 Db 669 GCTACTTCTGCTCAAGTTATGATCA-----AACACAGCT 634
 QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGly 204
 Db 633 AGCTGTGTTTAAAGTCTGCGCGCTCATATCAGGTTTGAGCTCGCTTCAAGTTATGGA 574
 QY 205 ArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArg 224
 Db 573 TCAAGCACT-----ACTGTCTCTAGTTTAAAC-----TCT 544
 QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly-----AlaVal 241
 Db 543 GCGCTGTGTGTTCCAGTTATGGCTCAAGCTCTACTGTGCGGTGTTTAAAGTCTGCCCTCA 484
 QY 242 AspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMet 261
 Db 443 AGTTACGATTTAAGCAGCGGACGC----- 460
 QY 262 LeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAsp 281

[illegible]

```

source
1..833
/organism="Vitis vinifera"
/cultivar="Ugni blanc"
/db_xref="taxon:29760"
/clone="PT005F08"
/clone_1db="Veraison Grape berries SuperscriptTM Plasmid library"
/dev_stage="Vernaison stage"
/site_1: SalI;
/site_2: NotI; Oriented library"
BASE COUNT      183 a      251 c      137 g      262 t
ORIGIN
Alignment Scores:
Pred. No.:       0.0757          Length:         833
Score:           111.50          Matches:        46
Percent Similarity: 45.77%      Conserved:     19
Best Local Similarity: 32.39%    Mismatches:   54
Query Match:      4.63%         Indels:        23
DB:               14            Gaps:          6

US-09-825-414-66 (1-487) x BQ800257 (1-833)
OY      336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyClyTyTyrAlaGlyVal 355
      ||| |||||:::||| ||||||||| :::: |||
Db      444 GGCGGGCTCTCCCTGCCTGCCTCACCTGGGTGGCTTCCTTCCTTGCGGCCAGCT 503
      |||
OY      356 SerLysLeuGlnMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
      |||
Db      504 TCC-----TTCGCTGGGGCTGCACCTCCGCC 530
      |||
OY      376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThrThraAcly 395
      ||| :::::::|||| | |||
Db      531 TCAGTAGTCTCTTTGGCTTCACACAGCGGCTTCGCC-----TCCACTGCTGA 584
      |||
OY      396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
      ::::: |||:::||| |||
Db      585 GNCCTC-----CCTTCGGTT-----TTCTCGCTTCGCTCCCTCAACC 623
      |||
OY      416 ThrAlaSerSerThrSerTyTyrAlaAlaAspGlnThrValLysLeuAlaLysThrVal 435
      |||||:::|||:::|||:::|||:::|||:::||| |||
Db      624 GCGGCTCTCTCAACAACCTCTTTGTCTCACTTAACCTCAACAGAGCTTCACCTTG 683
      |||
OY      436 LysAspMetSerGlyLysAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455
      ||||| ||| |||||:::||| |||
Db      684 GCTTACGCTTCAGGCGCGCTCCCTCCATCGAGGCTC-----TCCATTGTTTCT 737
      |||
OY      456 AsnLeuArgHisArgSerAlaProGluAlaAspIleGlu---GluGlyGlyIleSerAla 474
      ::::: ||||| :::: |||
Db      738 TCAGTAACCTGCCTCGAGACACCGGTGAGACCTCAACCTTTAGCGGTCCCTCCACC 797
      |||
OY      475 PheSer 476
      ::::::
Db      798 TGCTCA 803

RESULT 7
AL508002/c      680 bp      mRNA      linear      EST 04-JAN-2001
LOCUS          AL508002/c
DEFINITION     Hordeum vulgare Barke developing carypopsis (3.-15.DAP)
ACCESSION      AL508002
VERSION        AL508002.1 GI:12034217
KEYWORDS
SOURCE
ORGANISM       Hordeum vulgare.
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
                ; Triticeae; Hordeum.
                EST.
                Hordeum vulgare.
                Hordeum vulgare.
REFERENCE
AUTHORS        Michalek W., Weschke W., Pleissner K.-P. and Graner A.
TITLE          EST sequencing and analysis in barley
JOURNAL        Unpublished (2000)
COMMENT        Contact: Michalek W
```


QY	436	Lysaspmetserglylualia11easerfrrhglvalaserleuarSerhrvalsn	455
Db	666	GCTTACAGTTAAGGTGCGCGCTGCTTCCTTCACATGGAGCCTCT-TCCTTTGTTCT	719
QY	456	AsnLeuAgnHsArg-----SerAlaprogLuLa	465
Db	720	TCAGTAACGTGGCTGTGAGACAGCGGTGAGCGCTTACTCTTTTACACGCTCCCTTC	779
QY	466	AspIleGluGluGlyGlyIleSerAlapheser	476
Db	780	ACCTTGTCACATTCTCCCTCGTTTTCATTCATTC	812
RESULT 9			
LOCUS	CNS079ZS	1150 bp	DNA
DEFINITION	T3 end of clone XBB0AA002D09 of library XBB0AA from strain CBS 4732	linear	GSS 08-JUL-2000
ACCESSION	AL435950		
VERSION	AL435950.1	GI:12219363	
KEYWORDS	GSS.		
SOURCE	Pichia angusta.		
ORGANISM	Pichia angusta		
REFERENCE	Eukaryote, Fungi: Ascomycota; Saccharomycotina: Saccharomycetes; Saccharomycetales: Saccharomycetaceae; Pichia.		
AUTHORS	1 (bases 1 to 1150) Souciet, U.L., Aigle, M., Attiguenave, F., Blandin, G., Bolotin-Pukhara, M., Bon, E., Brothier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Llorente, B., Malpertuy, A., Neugeglise, C., Ozier-Kalogeropoulos, O., Potier, S., Saurin, M., Tekala, F., Toffano-Nloche, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies		
JOURNAL	FEMS Lett. 487 (1), 3-12 (2000)		
MEDLINE	20584711		
PUBMED	11152876		
REFERENCE	2 (bases 1 to 1150) Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Attiguenave, F. and Dujon, B.		
AUTHORS	Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta		
TITLE	FEMS Lett. 487 (1), 76-81 (2000)		
JOURNAL	20584723		
MEDLINE	11152888		
PUBMED	3 (bases 1 to 1150)		
REFERENCE	Genoscope.		
AUTHORS	Direct Submission		
TITLE	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
JOURNAL	This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.		
FEATURES	Location/Qualifiers		
source	1..1150 /organism="Pichia angusta" /strain="CBS 4732" /db_xref="taxon:4905" /clone="XBB0AA002D09" /clone_11b="XBB0AA" /note="end : T3"		
BASE COUNT	250 a 352 c 258 g 287 t 3 others		
ORIGIN			

Pred. No. :	0.434	Length:	1150
Score:	107.00	Matches:	63
Percent Similarity:	42.70%	Conservative:	57
Best Local Similarity:	22.42%	Mismatches:	107
Query Match:	4.45%	Indels:	54
DB:	17	Gaps:	13

US-09-825-414-66 (1-487) x CNS07925 (1-1150)
<p>OY 216 ThrPheSerAlaLeuAsnValValaArgThrValLeuAlaProValaLeuAlaSerArgPro 235</p> <p>DB 114 ACTGCTTCTAGATTAATTAAGGCCAGCGTCAGCGATCT-----AGCAGTACAGG 164</p> <p>OY 236 SerValGlnGlyAlaValaAspPheGlyValSer-----Thr 247</p> <p>DB 165 GCATCCAGAGGGGCATCTTCGCGTCCACAGCTGCTTCGCTACGCAGATTACT 224</p> <p>OY 248 AlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArg 267</p> <p>DB 225 GCAACGACGACCGCTTCCACCTATGCTTACCTGTTCATATCCACATCAGGATAC 284</p> <p>OY 268 AspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAla 287</p> <p>DB 285 AATTCA---AGGGCTTTAGCCCACTGCCGA-----GCGTCAAGTCA 326</p> <p>OY 288 LeuSerGlnGluThrAspTyrLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer 307</p> <p>DB 327 ATTTCACGCAAAAC-----AAGCATTC 350</p> <p>OY 308 GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAsp 327</p> <p>DB 351 AGCACTACTTGTCCAGAGGTTC-----AGCTCAATGGCTGTCTTACCGCTTCTGGC 404</p> <p>OY 328 GlyLeuLysAlaValaArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla 347</p> <p>DB 405 AGTGCCCTAGAGGAGCTAGTGAATTTCCAGCTTTCTGCTGCAGCCTCGATATGTT 464</p> <p>OY 348 LeuAlaGlyGlyTyrAlaGlyLysSerLysGlnLysMetAlaThrLysAsnIleThr 367</p> <p>DB 465 TTCAC-----GCTGCCCGCCAGCTCCATTTCCAAACAGCGCTTCAAGCAATCAG 515</p> <p>OY 368 AspSerAlaThrLysAlaAlaValaSerGlnLeuSerAsnLeuValaGlySerValaGlyVal 387</p> <p>DB 516 ACGGTAGGATTCATCTAGTCTGCTCTAGCGCTCCAGCTCCAACTCCAACTCTCTAG 575</p> <p>OY 388 PheAla-----GlyTyrPheThrLysAlaGlyLeuAlaThrAspProAlaValLysAla 405</p> <p>DB 576 ACTGCTTCGAGCGGCTTCTCCAGCTCCACGCTCCAGCGCAAGCTCT-----CGCGTGGC 626</p> <p>OY 406 GlnSerPheIleGlnAspLys-----ValLysSerThrAlaSerSerThr 420</p> <p>DB 627 GACACCTTCTCCAGACGACCAACGCGTCTATGTCTCGAGCAACAGCTCTCCAGTGT 686</p> <p>OY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440</p> <p>DB 687 TCTACTTATGTTCCGACGACGCC-----AGCGTCACTCTAGTCAGAGT 731</p> <p>OY 441 GlnAlaIleSerSerThrGlyLysSerLeuArgSer-----ThrValaAsnAsnLeuArgHis 459</p> <p>DB 732 GGCCTTACGAAATATCCAGTTATCTTCCTGCTACGTGCACCATCAGCTCTGCTGAACA 791</p> <p>OY 460 ArgSerAlaProGlnAlaAspIleGlnGlyLysSerAlaPheSerArgSerGlu 479</p> <p>DB 792 GCATTTTCAGGCAAGTAGTGCATCCAAATGTGGAGTCAAGC-----TCCGAA 839</p> <p>OY 480 Thr 480</p> <p>DB 840 ACC 842</p>

LOCUS	AY107053	1763 bp	mrna	linear	HTC 25-MAY-2002
DESCRIPTION	Zea mays PC0082530 mRNA sequence.				

ACCESSION	AY107053
VERSION	AY107053.1 GI:21210131
KEYWORDS	HTC.
SOURCE	Zea mays.
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 1763) Halney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V. Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes Unpublished (2002) 2 (bases 1 to 1763) Coe,E.C. Direct Submission Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
JOURNAL	location/qualifiers
TITLE	1..1763 /organism="Zea mays" /db_xref="Maizedb:634619" /db_xref="taxon:4577" /clone="PCO082530" /clone.lib="Maize Mapping Project/Dupont Consensus Library"
FEATURES	/note=""this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of Bacs in conjunction with the Maize Mapping Project"
BASE COUNT	341 a 543 c 535 g 344 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0.913 Length: 1763
Score:	107.00 Matches: 127
Percent Similarity:	34.14% Conservative: 56
Best Local Similarity:	23.69% Mismatches: 189
Query Match:	4.45% Indels: 166
DB:	Gaps: 25
US-09-825-414-66 (1-487) x AY107053 (1-1763)	
OY	19 ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerThr---Thr 37 : Db 217 AGGACTACTTGCACGAGCGCTTCGGTGCCGCCGCGGTGCCGGCGGCGATCCAC 276
OY	38 SerCysArgAspLeuGlnAlaIleThrAspTryLeuLysHisValPheAlaIahis 57 : Db 277 TCTCTCCGCTGCAGCCGCT--- 294
OY	58 ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisnGluGln 77 : Db 295 -----CCAAGCTCCCA----- 306
OY	78 IleasPalalaLeuValGluThrArgAlaAsnArgLeuTYrSerGluGlyGluThrProAla 97 Db 307 -----GCNACGACGCCGACGCCGCTTCGTTCTCCGGTGGAGAAGCCCTGCGCC 351
OY	98 ThrIlealagluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThr----- 114 ::: :: :: : Db 352 GCTCAGCGCACCTCTGGGCGCGATTATGCGGGCGCGGCCACGCGTCCGCGCGTCA 411
OY	115 -----ThrAlaSerSerAlaPheGluAsnThrProPheAla 126 : :: : : :: Db 412 TTACGAGCGTAGCCCTGCACCTCCACGTCATCCCATCGATGCCAAGAGCGTCCGCTTCAGT 471
OY	127 AlaAlaSerValLeuGlnTyrmetylProAlaIleAlanLysGlyAspTryPleuAlaThr 146 Db 472 GCCACGCTCTTCCCCTTCACAGCGCCACACCATCTCTCGCTCC-----TTGCTTACA 522

Oy	147	ProLeuLysProLeuThrProLeuLeuSerGlyAlaLeuSerGlyAlaMetAspLeuVal	166
Db	523	CCCCCTGACCATGCAAGAAAGGCCGACAAAGGAGATGACTGGGACAGCTGCATATCC	582
Oy	167	GlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAsp	186
Db	583	CGGGTGTGTGTCSCATACCGCAAGTCTGTGTCTCCGAGAGTGTGCTTACCTCG---AAC	639
Oy	187	LysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGln	206
Db	640	AAACTCTTCCACCAAGCATTTATC-----GCCAAGGCGCGCAG	678
Oy	207	ValValAspMet---GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr	225
Db	679	ATCATCAACGCTACAGCGCTTTTCTGTGAACAGTTTCGACGCTTGAGACCGGTGGC	738
Oy	226	Val-----LeuAlaProAlaLeuAlaSerArgPro	235
Db	739	GCTGCCCTGAGATGGCAAGCTGCTCCGGGGTTCCACCGGTACCATTTGGCCCA	798
Oy	236	---SerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGly-----	250
Db	799	CTCATGTGCGAGAAACTCATGTTGACGAAGTGGAGAAAGAAAGATCGGGCTCCCC	858
Oy	251	-----LeuValAlaAsnAlaGlyPheGlyAspArg	260
Db	859	GTTGCGCTGGCTTACGAGCAGCGACGGCGGTGTGTGTGTACTGCTGGCTTGGCAACCC	918
Oy	261	MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla-----	274
Db	919	-----TGCCCGCTGAGCCATGACACGATCCAGGAGATGCGCGGCTGAGCGCAGC	972
Oy	275	-----PheValLeuGlyMetLys-----AspLysGluProLysAlaAla	287
Db	973	AACTGTGGGTTCCTTTGGTGTCTCAAAACCAACAGCGTGGATAGGAGACAGACGCCGTG	1032
Oy	288	LeuSerGlnIuThrAspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db	1033	CTGACGGATTT-----GCTCGGCGAGGATTCCTGAGAGCGAGT-----	1070
Oy	308	GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAsp	327
Db	1071	-----CGAGGCTGCGCG-----CCCTGCGACCAAGCGCTGGT	1103
Oy	328	GlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla	347
Db	1104	GGA-----CAGAGAGCCCTCT	1121
Oy	348	LeuAlaGlyTyrAlaGlyAlaSerLysLeuGlnLysMetAlaThrLysAsnIleThr	367
Db	1122	GAACACACCGCGCATGGGCTGTTCTTAAGCAACACGCGGTG-----	1163
Oy	368	Asp-SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal-Gly-----	383
Db	1164	GAACCTCGTACCGAGCGCGCTCCGCCAGCTGCCGTGCTGGCGTGGCCCCCGCGCGG	1223
Oy	384	-----SerValGlyValPheAlaGlyTrpThr	393
Db	1224	CGACACCGCGGTGACGCCATGTTAACCGGTGACCGGTGGGTGGGATGTGATGAGACA	1283
Oy	393	hr---AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlnSerPheIleGlnAspL	412
Db	1284	CTGAGCTGGTGAGC-GGGAGAGACTGCTGTGATCACCGGGGTGAG-----ATCGCAGAGA	1366
Oy	412	ysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuA	432
Db	1337	AAAGTAAAGAGGTGATGTC-----GACCAACCGGTACAG---G	1372
Oy	432	lalysThrValLys-----AspMetSerGlyAlaIleSerSerThrGlyAlaSerL	450
Db	1373	CAGAGCAGCAGAGAGACCGCGGAGGAAGCGGCCAAGGACAGTTTCAAGAGGTGGCACACACT	1432

```

Oy 450 euargSerThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGln 470
    |||||
Db 1433 ACCGAGCATGGAGAGATTATTTCACGGCTCAAGCAACCGCTGCTTATTCGTC 1492
Oy 470 lylGlyIleSerAlaPheSerArgSerGluThrProPheGlnLeu 484
    |||||
Db 1493 GTGGGGTTCCTGATTTCACAGATTGTGTCCTCCCTACTCCCTCA 1536

RESULT 11
BO800379
LOCUS BO800379 841 bp mRNA linear EST 30-JUL-2002
DEFINITION EST 7414 Veralson Grape berries SuperScriptTM Plasmid Library Vitis
vinalifera cDNA clone PT009E05 3', mRNA sequence.
ACCESSION BO800379
VERSION BO800379.1 GI:22015345
KEYWORDS EST.
SOURCE Vitis vinifera.
ORGANISM Vitis vinifera.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 841)
AUTHORS Abdal,P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F.,
Couture,C., Dedaldechamp,F., Delrot,S., Glissant,D., Grimplet,J.,
Hamdi,S., Romieu,C. and Terrier,N.
Handl, S., Romieu, C. and Terrier, N.
Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
or seeds) at Various Developmental Stages
Unpublished (2002)
TITLE
JOURNAL
COMMENT Contact: Delrot S.
TRANSPORT DES ASSIMILATS
UMR CRNS 6161, Universite de Poitiers
Batiment Botanique, 40 Avenue du Recteur Pineau, 86022 Poitiers
Cedex, France
Tel: 00-33-(0)5-49-45-41-85
Fax: 00-33-(0)5-49-45-41-86
Email: Serge.Delrot@univ-poitiers.fr
Seq primer: SP6.

FEATURES
source
location/Qualifiers
1..841
/organism="Vitis vinifera"
/cultivar="Ugni blanc"
/db_xref="taxon:29760"
/clone="PT009E05"
/clone_lib="Veralson Grape berries SuperScriptTM Plasmid
library"
/dev_stage="Veralson stage"
/note="Organ: Fruit; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; Oriented library"
BASE COUNT 187 a 251 c 138 g 265 t
ORIGIN

Alignment Scores:
Pred. No.: 0.373 Length: 841
Score: 105.50 Matches: 48
Percent Similarity: 44.08% Conservative: 19
Best Local Similarity: 31.58% Mismatches: 56
Query Match: 4.38% Indels: 29
DB: 14 Gaps: 7

US-09-825-414-66 (1-487) x BO800379 (1-841)
Oy 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTYrAlaGlyVal 355
    |||
Db 444 GTGCTGGCTCTCTCTGCTTCACCTGCTTGGCTTCTTCTTCTTCCCTGGCGCAGT 503
Oy 356 SerLysLeuGlnLysMethAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
    |||
Db 504 TCC-----TCTGCTGGGGCTGCACCTCCGCC 530

Oy 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrPheThrAlaGly 395
    |||
Db 531 TCAGTAGTCTCTTCTTCTTGGCTTTCACACGCGGCTTCTGCC-----TCCACTGCTGGA 584

```

```

Oy 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
    |||||
Db 585 GTCTCT-----CTTCGGTT-----TTCTGCTTCCTCCCTCCACACC 623
Oy 416 ThrAlaSerSerThrThrSerTyValAlaAspGlnThrValLysLeuAlaLysThrVal 435
    |||||
Db 624 GCGGCTTCCTTACACACCTCTTGTCTCCACCTCACTCAACTCAACAGACTTCACGCTTG 683
Oy 436 LysAspMetSerGlyAlaAlaIleSerThrGlyAlaSerLeuArgSerThrValAsn 455
    |||||
Db 684 GCTTCAGCTTCAGGTCGCCCTCCCTCCCTCCACCTGAGACCTCT-----TCCTTTGTTCCT 737
Oy 456 AsnLeuArgHisArgSerAlaProGluAlaAspIleGluGlnGlyGlyIleSer--Ala 474
    |||||
Db 738 TCAGTACTGCGCTCGAGCA-----GCCGGTGAGACCTCTTACTCTT 779
Oy 475 PheSerArgSerGluThrProPheGlnLeuArgArg 486
    |||||
Db 780 TTTACGGCTCCTCTTCACCTGTTCATTCCTCGG 815

RESULT 12
BF265543/c
LOCUS BF265543 850 bp mRNA linear EST 23-OCT-2001
DEFINITION HV.CEA0012J12f Hordeum vulgare seedling green leaf EST library
HVCDNA0004 (Blumeria challenged) Hordeum vulgare cDNA clone
HV.CEA0012J12f, mRNA sequence.
ACCESSION BF265543
VERSION BF265543.2 GI:13262162
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 850)
AUTHORS Wing,R., Close,T.J., Kleinhof,A., Wise,R., Wei,F., Begum,D.,
Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi
D.W., Fenton,R.D., Oates,R. and Malo,D.
Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected incompatible (Mla13)
seedling leaf cDNA library
Unpublished (2001)
JOURNAL On Nov 17, 2000 this sequence version replaced gi.11196537.
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total bp bases = 429
Seq primer: AATTACCTTCACTRAAGCG
High quality sequence stop: 743.

FEATURES
source
location/Qualifiers
1..850
/organism="Hordeum vulgare"
/cultivar="C116155 (Mla13)"
/db_xref="taxon:4513"
/clone="HV.CEA0012J12f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDNA0004 (Blumeria challenged)"
/tissue_type="seedling green leaf"
/lab_host="TJc121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
C.I. 16135 (Mla13) plants were greenhouse grown in the R
wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were challenged with isolate A27 (AVRMA13
) of Blumeria graminis f. sp. hordei, and leaves were
harvested 20 and 24 hr post-inoculation and snap frozen;
uninoculated leaves were harvested 20 hr post-inoculation
(Wei, Wise). In the TJ Close lab at the University of
California, Riverside, total RNA was prepared from each
sample pool, equal quantities of all three RNA pools were

```

combined, poly(A) RNA was purified from the mixture, one cDNA library was made, and 1 million pfu were in vivo excised to give plasmid SK(-) cDNA phagemids (Choi, Close). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

<http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9papes/dgn/31/cover.html>)"

BASE COUNT 139 a 286 c 294 g 131 t
ORIGIN

Alignment Scores:

Pred. No.: 0.38 Length: 850
Score: 105.50 Matches: 59
Percent Similarity: 36.47% Conservative: 34
Best Local Similarity: 23.14% Mismatches: 100
Query Match: 4.38% Indels: 62
DB: 12 Gaps: 9

US-09-825-414-66 (1-487) x BF265543 (1-850)

QY 11 Proglyalaalamegluserpheargthralsasrpaalaserleualseaser 30
|||||

Db 670 CCGGGGGTCCGCGGAGAGCC-----AAGTGCTACCAAGCGCCAGCCGGTAAATAT 617
|||||

QY 31 Valargservalserthrthsercysargaspueuglnalaletthraspyrleuuls 50
|||||

Db 616 GCTGAGGCGCTCACCAGAGGGTCCGCCACACAGCGTAGCCCTGCATCAAGCCCATAG 557
|||||

QY 51 Hishisvalphealaalalhisargpservalleuglyserproaspqluargspala 70
|||||

Db 556 CACAC-----CTGCGATCCCGCAGATGGAGGATCT 524
|||||

QY 71 Alaleualahisangluginleaspaaleuvalgiuthrargalaasnarqleuylr 90
|||||

Db 523 CTGGGCGTGCACACAGTGGGGTCCAGATGCGAGGCCCTCCGAGCAATCTG 464
|||||

QY 91 Sercluglyluthrproalathrlealagluthrphaalalyalaaglulspheasp 110
|||||

Db 463 CGCCCGGCGCTTGAACAGACGCTCCGAACTGATGCGCTTGGACAG----- 410
|||||

QY 111 Argleualathrthralsasrseralaphhegluasnthrprophaalalalaser 130
|||||

Db 409 GATTCGGGGACACAGCCCGAGGGCTCCGAGCTCCCA-----CCGCGAGTG 359
|||||

QY 131 Leuglntyrmetglinproalaileasnlyslasprtleualathrproleuyls 150
|||||

Db 358 GATCACCTCCAGCTCCCTGCTTGGCGAGAGTCTCTGGCAGCGCAATCCCGCGT 299
|||||

QY 151 Leuthrproleuileserglyalaleuuserglyalameaspelinalglythrlismet 170
|||||

Db 298 GTCCACCCGATGTCGCGCGGAGTCCCGCT-----CAGGTGCGAGCGGGCTG 248
|||||

QY 171 Metaspargalalargglyasplleuistyrleuserthrserproaspalsleuistasp 190
|||||

Db 247 CTCGGAGACGGCCGAGCTTGGAGCGGTACCGCAT----- 200
|||||

QY 191 Alamelalavalservalysarghisserproalaleuglyarglnalvalasapket 210
|||||

Db 199 -----GCTCGGGGGCC----- 188
|||||

QY 211 Glyllealavalglnthrphaseralaleuasn-valvalargthrvalleu----- 227
|||||

Db 187 -----GCTCTTGACGGTCCGGCCAGCGGTGATGCGGCCGCC 152

QY 228 -----Alaproaleualasrargproservalglnalvalal----- 241
|||||

Db 151 GAATGCGCGGACGCTGCGGACGAGGTATCCGCGCGGGCGGTCCGAGGAAGCTGT 92
|||||

QY 242 -----Asppeglyvalserthrthraspyrleuuls 250
|||||

Db 91 CGTCTGCTGACGCGCCAGCGGCGCATTTGCTGTGGCTGG 49
|||||

RESULT 13

CNS076CP/ 888 bp DNA linear GSS 07-JUL-2001

LOCUS T7 end of clone BB0A002G01 of library BB0A from strain CBS 4732

DEFINITION of Pichia angusta, genomic survey sequence.

ACCESSION

VERSION

KEYWORDS

SSS.

AL431231.1 GI:12214643

SOURCE

ORGANISM

Pichia angusta

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE

AUTHORS

Souchet, J.L., Aigle, M., Artiguenave, F., Blandin, G., Bojotin-Pukhara, M., Bon, E., Brothier, P., Casaregola, S., Del-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Florente, B., Malpertuy, A., Neuvéglise, C., Ozier-Kalodopolos, O., Portier, S., Saurin, W., Tekala, F., Toffano-Nioche, C., Weslowski-Louvel, M., Wincker, P., and Weissenbach, J.

Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

FEBS Lett. 487 (1), 3-12 (2000)

TITLE

JOURNAL

MEDLINE

PUBMED

2 (bases 1 to 888)

Blandin, G., Florente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B.

Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta

FEBS Lett. 487 (1), 76-81 (2000)

20584723

11152888

3 (bases 1 to 888)

Genoscope.

Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

SOURCE

1. 888

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

Location/Qualifiers

```

Best Local Similarity: 23.05%      Mismatches: 119
Query Match: 4.38%                Indels: 77
DB: 17                             Gaps: 14

US-09-825-414-66 (1-487) x CNS076CP (1-888)

QY 214 ValGlnThrPheSerAlaLeuAsnValValArgThrValLeu-----AlaProAlaLeu 231
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 882 CTrACAGCATCTCTCGAATCCCACTCTACCCAGACGATTCATCTGCCCTMCCCT 823
QY 232 AlaSerArgProSerValGlnGlyAlaValAlaAspHeGlyValSerThrAlaGlyLeu 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 822 GCTTCA---TCGTCGCAACAAGCTCATTTGGCGAAGCGTTTCCAGTCTGGAACAC 766
QY 252 ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerVal---GlnSerArgAspGlnLeu 270
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 765 TCTGCCAGAGCC-----TCCGCTACGGCGGGAATCTCAATCCACCACGACGAT 715
QY 271 ArgGlyGlyAlaPheValLeuGlyMetLysAspLysGlyProLysAlaAlaLeuSerGlu 290
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 714 GGCTCCACGCTCTTAGTATTTATGSCCAG-----TCTCAGTCCGATATTAGC 667
QY 291 GlnThr-----AspTyrLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer 307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 666 AGTACAGGGGCATCCAGTTGGGCATCCATGCTCGCTCC---AGCAGTCTCTCTCGCT 610
QY 308 Gly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAla 325
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 609 ACCGAGATTACTGCACAGCAGCACCGCTTCACCTATGCTTACGCTCTTCATATATCC 550
QY 326 ThrAspGlyLeu-----LysAlaValArg 333
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 549 ACATCAGATACATTCACAGGGGCTTTAGCGCACTGCCGAGCCGTCACGTCGATTTCC 490
QY 334 SerLeuValSerAlaThrSerLeuThr---LysAsnGlyLeu----- 346
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 489 ACCCAAAACACGACATCCCGACGACTTGTGCAACGCTTCAAGCTCAATGCTCTCTCT 430
QY 347 -----AlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAla 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 ACCGCTTTCGCAAGTCTCTCTAGTGGAGCTAGTGGAAATTCAGTCTTCTGCTGCGAC 370
QY 363 ThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 TCGAGTAATRTTTTTCAGTCTGCTGCCACCTCATTTTCAACAGCCCTTCAAGCCAAAGTC 310
QY 383 GlySerValGlyValPheAlaGlyTyrPheThrAlaGlyLeuAlaThrAspProAlaVal 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 309 ACGACGCTAGTTCATCTAGTCTGCTGCTGCTAGCGCTTCCAGTGCACATTCATTCCT 250
QY 403 LysLysAlaGlyLysPheIleGlnAspLysVal----- 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 249 AAGACTCTTCGAGCGGCTTCTCCAGCTCCCTCAGCGCAACGTCGCCGTCGGACAGC 190
QY 414 -----LysSerThrAlaSerSerThrThrSer 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 189 TTCTCCAGACAGCAACAGGCTATTGCTGAGACACACAGCTTCACGATGTTTCTAGC 130
QY 423 TyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyAlaVal 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 129 TATGTTTCCGACAGCC-----ACGCTCACATCTAGCTCAGTGGCTGCTCT 85
QY 443 IleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAla 462
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 84 ACGAAATAACCACTTTATCTTG----- 61
QY 463 ProGluAlaAspIleGluGly 470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 60 CCGAGTGCACCATCAGCTCTGCT 37
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
BQ795937/C LOCUS BQ795937 579 bp mRNA linear EST 30-JUL-2002

```

```

DEFINITION EST 4875 Ripening Grape berries Lambda Zap II Library Vitis
viñifera cDNA clone RT022609 3', mRNA sequence.
ACCESSION BQ795937
VERSION BQ795937.1 GI:22010903
KEYWORDS EST.
SOURCE Vitis vinifera.
ORGANISM Vitis vinifera.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 579)
AUTHORS Abbal,P., Agase,A., Ageorges,A., Atanasova,R., Barrieu,F.,
Couture,C., Dedaldecamp,F., Delrot,S., Gleysant,D., Grimplet,J.,
Hamdi,S., Romieu,C. and Terrier,N.
Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
or seeds) at Various Developmental Stages
JOURNAL Unpublished (2002)
COMMENT Contact: Romieu C.
Unité de Recherche des Produits de la Vigne
Institut National de la Recherche Agronomique
2, place Viala, 34 060 Montpellier Cedex 01, France
Tel: 00-33-(0)4-99-61-28-62
Fax: 00-33-(0)4-99-61-28-57
Email: romieu@ensam.inra.fr
Seq primer: 77.

FEATURES
     source          location/Qualifiers
     source          1..579
                     /organism="Vitis vinifera"
                     /cultivar="Shiraz"
                     /db_xref="taxon:29760"
                     /clone="RT022609"
                     /clone.lib="Ripening Grape berries Lambda Zap II Library"
                     /dev_stage="ripening stage"
                     /note="Organ: Fruit; Vector: Lambda Zap II; Site:1: Eco RI
                     /note:2: XhoI; Oriented library, construction described
                     in Generation of ESTs from grape Berry (skin, pulp or
                     seeds) at various developmental stages by Terrier,N.,
                     Ageorges,A., Abbal,P., Romieu,C. In J. Plant Physiol. 158
                     (12): 1575-83 2001"
BASE COUNT 177 a 136 c 174 g 92 t
ORIGIN
Alignment Scores:
Pred. No.: 0.222      Length: 579
Score: 105.00      Matches: 45
Percent Similarity: 43.71%      Conservative: 21
Best Local Similarity: 29.80%      Mismatches: 53
Query Match: 4.36%      Indels: 32
DB: 14              Gaps: 6

US-09-825-414-66 (1-487) x BQ795937 (1-579)

QY 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyVal 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 528 GTGCTGCTCTCTCTCTCTCTCAGCTGATGGTTCCTTCTCTGCGCGCAGCT 469
QY 356 SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 468 TCC-----TTCGCTGGGCTCGACGTCGCTCC 442
QY 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrPheThrAlaGly 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 441 TCGATGCTCTCTCTTGGCCCTTCTCAACAGCGGCTTCGCC-----TCCACTGCTGA 388
QY 396 LeuAlaThrAspProAlaValLysLysAlaGlySerPheIleGlnAspLysValLysSer 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 387 GTCCTP-----CCTTCGCT-----TTCCTCTCTCTGCTCTCTCAACC 349
QY 416 ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrVal 435
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 GCGGCTTCTCTTCAACAGCTCTTGTGCTCAACTTCAACAGACGATTCAGCTTGG 289
QY 436 LysAspMetSerGlyLunAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455

```

```

Db 288 GCTTCAGCTTCAGTCCGCTCCCTCCACTGAGCCTCT-----TCCTTGTTCCT 235
QY 456 AsnLeuArgHisArg-----SerAlaProGlnAla 465
Db 234 TCAGTAAGTGGCTGTGGAGACCGGTGGAGCCTTACCTCTTTTACGGCTCCCTCTTCC 175
QY 466 AspIleGluGluGlyGlyIleSerAlaPheSer 476
Db 174 ACCTGTTCAATTCCTCGGTCTTTTCAATTCCTCC 142

RESULT 15
BM077210 806 bp mRNA linear EST 05-FEB-2002
LOCUS TREST-A4136 TREST-A Hypocrea jecorina cDNA clone Tr-A4136 5', mRNA
DEFINITION
ACCESSION BM077210.1 GI:18498392
VERSION BM077210.1
KEYWORDS EST.
SOURCE Hypocrea jecorina.
ORGANISM Hypocrea jecorina
Eukaryota; Fungi; Ascomycota; Pezilomycotina; Sordariomycetes;
Hypocreales; Hypocreaceae; Hypocrea.
REFERENCE
1 (bases 1 to 806)
Chamberg, F.S., Bonaccorsi, E.D., Ferreira, A.J.S., Ramos, A.S.P.,
Ferreira, J.R.Jr., Abramo-Neto, J., Parah, J.P.S. and El-Dorri, H.
Elucidation of the metabolic fate of glucose in the filamentous
fungus Trichoderma reesei using expressed sequence tag (EST)
analysis and cDNA microarrays
J. Biol. Chem. 277 (16), 13983-13988 (2002)
JOURNAL MEDLINE
COMMENT
21950703
Contact: El-Dorri, Hamza
Department of Biochemistry
Institute of Chemistry, University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
BRASIL
Tel: (55) 11-38183848
Fax: (55) 11-38183848
Email: dorri@iq.usp.br
PCR Primers
FORWARD: Universal M13 forward primer
BACKWARD: Universal M13 reverse primer
Plate: 44 row: A column: 8
Seq primer: M13 reverse primer
High quality sequence stop: 806
POLYA=No.
FEATURES
source Location/Qualifiers
1..806
/organism="Hypocrea jecorina"
/strain="QM9414 (ATCC26921)"
/db_xref="taxon:51453"
/clone="Tr-A4136"
/clone_1lb="TREST-A"
/sex="asexual"
/tissue_type="mycelia"
/dev_stage="18 hr glycerol-grown culture"
/lab_host="E. coli SOLR cells (kanamycin resistant)"
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:
XhoI; anamorph=Trichoderma reesei; Cloned unidirectionally
5' end of the cDNA cloned into EcoRI site of
pBluescript. Primer: Oligo (dT). Average insert size: 1,2
kb; Uni-ZAP XR Vector system -5' adaptor sequence:
5'GTAATTCGCGACGAG3' -3' adaptor sequence:
5'CTCGAGTCTTTTCTTTTCTTTT3'."
BASE COUNT 171 a 289 c 198 g 148 t
ORIGIN

```

Alignment Scores:

Pred. No.:	0.395	Length:	806
Score:	105.00	Matches:	61
Percent Similarity:	37.73%	Conservative:	22
Best Local Similarity:	27.73%	Mismatches:	70
Query Match:	4.368	Indels:	67

```

DB: 13 Gaps: 10
US-09-825-414-66 (1-487) x BM077210 (1-806)
QY 301 IleIysSerAlaSerTyrSerIleValAlaLeuAsnAlaGlyLys----- 315
Db 171 TTGAAAACGTTGGCCCTGCTCGGCCCA6CGCAACGTCGCAAGTCCTCCGCC 230
QY 316 -----ArgMetAlaGlyLeuProLeuAsp 323
Db 231 CTCCTCGCCGCCACAAATTCACCGTCACCGTCCTCCGCCGCCCTCTCCCATCC 290
QY 324 VALAlaThrAspGlyLeuLysAlaValArg---SerLeuValSerAlaThrSerLeuThr 342
Db 291 ACCTTCCCGAGCGGCGTCCGCTCGACGTCGACTTACGCTCGAGACCTCGAGCCTGACG 350
QY 343 LysAsnGlyLeuAlaLeuAlaGlyIleYrAlaGlyValSerLysLeuGlnLysMetAla 362
Db 351 -----GCCGCCCTTCCGCCCAAGACCGCTGCTCCACCGTCGCGACGCCGCC 401
QY 363 ThrLysAsn-----IleThrAspSerAlaThrLysAlaAlaValSerGlnLeu 378
Db 402 CTCAGAGACGACAAAGAGCTCATTCAGCGCCGCCGCGCGCTCAAGCCTTC 461
QY 379 -----SerAsnLeuValGlySerValGlyValPhe 388
Db 462 CTGCCCTCGAGTGGCTGGCGACCTGACCAAGAGAGCTCCGCCCAAGCTGCCCTTTC 521
QY 389 AlaGlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysAlaGlu----- 406
Db 522 GCA-----ACCAAGCTCGAGATCACG 542
QY 407 SerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAla-- 425
Db 543 CGGTACCTCGAGACAGCCCAAGACGACGCTCTACGTCNACCGCTGTACAGCGGG 602
QY 426 -----AspGlnThrValLysLeuAlaIsthrValLysAspMetSerG1 440
Db 603 CCTTCTTCGACTGGGACTACAGTACAACTTCATCTTCAGAGGCTGATCAAGC-- 660
QY 440 yGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisAr 460
Db 661 ----CCGTGCTGACAGCGGGGCAACACCGCTTTTAACTTCGAACCTTCGAAGCG 716
QY 460 gSerAla-----ProGluAlaAspIleGluGly 471
Db 717 TTCGCCAGGCGCGTGTGGGGTCTTTTAAACCGAGGAAGCAAGAAACGGCGGT 774

```

Search completed: January 31, 2003, 07:16:29
Job time: 1589.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 31, 2003, 04:12:39 : Search time 54.556 Seconds
(without alignments)
2737.582 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407

Sequence: 1 MHINQSAQPPGVAMSEFRT.....EEGISAFAFSEPFQRLRL 487

Scoring table:

BLOSUM62	
Xgapop 10.0, Ygapext 0.5	
Xgapop 10.0, Ygapext 0.5	
Xgapop 6.0, Ygapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n_model -DEV=xlp
-Q=/cgn2_1/USP10.spool/US09825414/rnact_28012003_161742_26114/app_query.fasta.1.1294
-DB=Issued_Patents_NA -QEMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=PCT -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MILLEN=0 -MAXLEN=2000000000
-USER=US09825414.@CGN.1.1.40.@runact_28012003_161742_26114 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/1/lna/PCUTS.COMB.seq:*
- 6: /cgn2_6/ptodata/1/lna/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	146.5	6.1	9542	4 US-08-968-685A-9	Sequence 9, Appl1
2	140	5.8	7100	4 US-09-308-375-1	Sequence 1, Appl1
3	132	5.5	3300	1 US-08-194-290-6	Sequence 6, Appl1
4	132	5.5	3300	2 US-08-614-377A-6	Sequence 6, Appl1
5	132	5.5	3300	4 US-09-142-648B-6	Sequence 6, Appl1
6	131.5	5.5	5393	2 US-08-591-079-9	Sequence 9, Appl1
7	130	5.4	46819	4 US-09-453-702B-72	Sequence 72, Appl1
8	127.5	5.3	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
9	126.5	5.3	32768	4 US-08-961-527-71	Sequence 71, Appl1
10	126.5	5.3	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
11	125.5	5.2	3666	2 US-08-682-517-13	Sequence 13, Appl1
12	125.5	5.2	3666	2 US-08-682-517-14	Sequence 14, Appl1

13	125.5	5.2	4197	2 US-08-682-517-7	Sequence 7, Appl1
14	125.5	5.2	4197	2 US-08-682-517-8	Sequence 8, Appl1
15	123	5.1	2489	3 US-09-141-047-7	Sequence 7, Appl1
16	122.5	5.1	3123	4 US-09-134-001C-2528	Sequence 2528, Ap
17	122	5.1	3979	4 US-09-085-199B-10	Sequence 10, Appl1
18	122	5.1	17612	3 US-08-911-853-29	Sequence 29, Appl1
19	122	5.1	17612	4 US-09-479-409-29	Sequence 29, Appl1
20	122	5.1	17612	4 US-09-479-453-29	Sequence 29, Appl1
21	121	5.0	2350	4 US-08-961-527-34	Sequence 34, App
22	120.5	5.0	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
23	119.5	5.0	4483	4 US-08-961-527-363	Sequence 363, App
24	119	4.9	25165	4 US-09-453-702B-39	Sequence 39, Appl1
25	117.5	4.9	1464	4 US-08-809-326A-3	Sequence 3, Appl1
26	117.5	4.9	1947	4 US-08-809-326A-17	Sequence 17, Appl1
27	117	4.9	2742	3 US-08-911-853-16	Sequence 16, Appl1
28	117	4.9	2742	4 US-09-479-409-16	Sequence 16, Appl1
29	117	4.9	2742	4 US-09-479-453-16	Sequence 16, Appl1
30	117	4.9	2780	1 US-08-441-139-1	Sequence 1, Appl1
31	116.5	4.8	4724	4 US-09-066-046-3	Sequence 3, Appl1
32	115.5	4.8	3106	2 US-08-591-079-5	Sequence 5, Appl1
33	115.5	4.8	3695	4 US-09-453-702B-18	Sequence 18, Appl1
34	115.5	4.8	6744	1 US-08-119-125A-2	Sequence 2, Appl1
35	114	4.7	1436	4 US-08-961-527-365	Sequence 365, App
36	114	4.7	7766	4 US-09-125-619-3	Sequence 3, Appl1
37	113.5	4.7	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
38	113	4.7	1732	4 US-09-449-335-5	Sequence 5, Appl1
39	112.5	4.7	6270	1 US-08-418-893D-25	Sequence 25, Appl1
40	112.5	4.7	6790	1 US-08-418-893D-22	Sequence 22, Appl1
41	111.5	4.6	1702	4 US-08-961-527-367	Sequence 367, App
42	111.5	4.6	13842	4 US-09-105-537-30	Sequence 30, Appl1
43	111.5	4.6	22671	4 US-08-976-259-14	Sequence 14, Appl1
44	111.5	4.6	36778	4 US-09-105-537-5	Sequence 5, Appl1
45	111.5	4.6	38506	3 US-09-320-878-19	Sequence 19, Appl1

ALIGNMENTS

RESULT 1

US-08-968-685A-9

Sequence 9, Application US/08968685A

Patent No. 6214961

GENERAL INFORMATION:

APPLICANT: TUCKER, KENNETH

APPLICANT: PLOSILA, LAURA

TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE

TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: PENNIE & EDMONDS LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/968, 685A

FILING DATE: No. 6214981ember 12, 1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Baldwin, Geraldine F.

REGISTRATION NUMBER: 31,232

REFERENCE/DOCKET NUMBER: 7969-060

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

; LENGTH: 9542 base pairs
; type: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-968-685A-9

Alignment Scores:

Pred. No.: 9,51e-05 Length: 9542
Score: 146.50 Matches: 111
Percent Similarity: 36.00% Conservative: 69
Best Local Similarity: 22.20% Mismatches: 181
Query Match: 6.09% Indels: 139
Gaps: 23

US-09-825-414-66 (1-487) x US-08-968-685A-9 (1-9542)

QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHis 51
Db 1346 AAGACTTAACTATCAGAGGTGTCACAGCACAGCCATTAACCGAT----- 1393
QY 52 HisValPheAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
Db 1394 -----CATTAACATCGGTGGTGTACAAAATGGCGATGCTGAAGTTCAA 1438
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArg-----AlaAsnArg 88
Db 1439 CTTCCTGAACCTTTAACAGCGCTTAATATGTTACCACTGAAACCTAACCGCCACAGAG 1498
QY 89 LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys 108
Db 1499 AAGATTCCTGAGGCAAAACC----- 1519
QY 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla 128
Db 1520 -----CGCCTT---ACCACAGATAAATGTTTACCAATGAT-----ATG 1558
QY 129 SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeuAlaThrProLeu 148
Db 1559 AATGCAATGTGTAAGAACCAAACTTACTCTGATAAGACACTGGCATTCATGACAGGTGT 1618
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeu----- 159
Db 1619 CAAAAGATTACCAAACTTACTGCTGTGTAGTACATGACAGTCGCGCAACTTATGACAG 1678
QY 160 -----SerGlyAlaMetAspGlnValGlyThrLysMet 171
Db 1679 CTTAAAAAAGTTAACCAACCGCTGAAGAGTCTCTACAAACCTTACCCTTAAAAAGTGA 1738
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 1739 GATTAATAATGCTAATGAT-----GCTAATGACAGC 1768
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMet--- 210
Db 1769 AAAATTCATCCCGGTGTAATAAATACCAACCAAGCGTACTGCAATCAACCCCTAAAA 1828
QY 211 -----GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1829 CTCAAAGGTGAACCGGTGATGATTACAAC----- 1861
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal 245
Db 1862 -----GNAACAAATGCTACAGTTACCTTTGGGCTT 1891
QY 246 SerThrAlaGlyGlyLeu---ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerVal 264
Db 1892 AACCAAAATAACGGTGTGACCGCTTGCAACAGCACCTTAACACGATGGCTTATCTGTT 1951
QY 265 Gln-----SerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAsp 281
Db 1952 AAAAACACGATAGTACAAACAAATCCAAAGTCGCTGTGATGGCATTTACATTCATGAT 2011

QY 282 -----LysGluProLysAlaAlaLeuSerGluGluThr-----AspTyr 294
Db 2012 ATCAGCATAGTAAAGCAGAGGTGCTGCAATGAAATACCACTCCATTACAGAGACGGT 2071
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 2072 ATTT-----GGTTTGGCTAATATATCTGT 2095
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAsp-----GlyLeuLysAla 331
Db 2096 TCATTGGATGTGAACCAAAACCCCGCTTAACCCCACTGGCATTAAAGCAGGTGAAGAG 2155
QY 332 ValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGly 351
Db 2156 CTGACCAATATGTCAAATCGGCATTAAACCTGCTACCAATGCT----- 2197
QY 352 TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThr 371
Db 2198 ---GGGACAGCTAGACTTATGAACCGCTTAAGCACTGCTAATTAACGAAAATTCAGGCTCT 2254
QY 372 LysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyr 391
Db 2255 GCCGCCACCATTAAGACTTTATACACTTACACAAAGTACCGCTGACCTTGCAAGT--- 2311
QY 392 ThrThrAlaGlyLeuAlaThrAspProAlaVal---LysLysAlaGluSerPheIleGln 410
Db 2312 -----GATACAGCTCTAATGTACCACAAAAATCGCGAGATTTTAAAG 2356
QY 411 AspLys---ValLysSerThrAlaSerSerThrThr-----SerTyrValAla 425
Db 2357 GTTAAGGTGTGAACCAACCCAGCTGATGATTTAACCAAAATTAACATCGGTGTGCTGT 2416
QY 426 AspGln-----ThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 2417 GATAGTACCGATATAGCTTAAACCGTTAACTGCTAAACCTTAAAGGATCTTGAGCGC 2476
QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 2477 GTTAATACTTAACCCCTTAACCTGACGAGATTAAGATTACCTGTAAGACAGTGGCAACAC--- 2533
QY 461 SerAlaProGluAlaAspIleGluGluGlyLysIleSerAlaPheSerArgSerGluThr 480
Db 2534 -----ACCGCTAAGCTACAAATGTGTGATTAAAC---TTTAGCAACAAATAATACA 2581
RESULT 2
US-09-308-375-1
; Sequence 1, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394-PCF
; CURRENT APPLICATION NUMBER: US/09/308,375
; CURRENT FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4
; EARLIER FILING DATE: 1997-09-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7100
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-308-375-1
Alignment Scores:
Pred. No.: 0.000292 Length: 7100
Score: 140.00 Matches: 128
Percent Similarity: 33.44% Conservative: 81
Best Local Similarity: 20.48% Mismatches: 222
Query Match: 5.82% Indels: 194
Gaps: 26
US-09-825-414-66 (1-487) x US-09-308-375-1 (1-7100)

QY 15 MetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerVal 34
 Db 862 ATTGAATTATATAGCAGACAGCAAGTAATAGTTCAAACTTAATACACGGTATGCC 921
 QY 35 SerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuYsHisHisValPhe 54
 Db 922 AGTTCTATGGGCTCTAGTATATAGCAACAGCTGTTCAAGATATTGGAAAT 969
 QY 55 AlaAlaHisArgPheSerVal---IleGlySerProAspGluArgAspAlaAlaLeuAla 73
 Db 970 GCGATAAATAGTCTTATATGACACTGGAAAGCAATATATCAAGATCACAAATTCAAAGC 1029
 QY 74 HisAsnGluGlnIleAspAlaLeu-----ValGluThrArgAlaAsnArgLeuTyr 90
 Db 1030 TTGAATATGCAATTTTAGAGATTAAGCTCCACAGCTCAAAACAGCTCAATTCAGCCTCT 1089
 QY 91 SerGluGlyGluThrProAlaThrIleAlaGluThrPhe----- 103
 Db 1090 TCTTTTGA-----GCAGAACTAACCCCAACCTTCAAAAGCATGTCACCTATTTA 1140
 QY 104 ----- 106
 Db 1141 ATCTCGGTTCTTATTTCTACGGAGCTATCTGCACTTAAGAAATGCTATCCAGGCA 1200
 QY 107 GluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla 126
 Db 1201 ATGAAATGATATCTCTCATGACAAATATTTGCCGCTTATGAAAGACGGATTTATAAA 1260
 QY 127 AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThr 146
 Db 1261 TATATGAACTCTC-----CAAGATCTATGACTAGTATGATACACTTCAAT 1311
 QY 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
 Db 1312 AAATATCACAGATTTCTTCAAAATGACAGCGGATTTGGGAATGGGTTTCGATGAAGT 1371
 QY 167 GlyThrLysMetLeuAspArgAlaArgGlyAspLeuHisTyrLeuSer---ThrSerPro 185
 Db 1372 GAGCTCTCCACGTTAACGAAACTGCCCAAGTTCTTCAAAATGTCCTGATTTAACTCCC 1431
 QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
 Db 1432 GATGATACAGTTTACACTCTAACGGCAGCAAGTCTCACTTAATATTCAGCAAAATGAT 1491
 QY 206 GlnVal-----ValAspMetGlyIleAlaValGlnThrPhe 217
 Db 1492 TCATATTCATTCGACATTAATTAATGAAGTTGATTAATTAATCTGCTTACCACTCTA 1551
 QY 218 SerAlaLeuAsnValAlaArgThrValLeuAlaProAlaLeuAlaSerArgProSerVal 237
 Db 1552 GATCTGCCCAATTTCTATCCGTAAA-----GCTGGTTCAACTGCTTTCATTCGGGGTA 1605
 QY 238 GlnGlyAlaValAspPheGlyValSerThrAla-----Gly 249
 Db 1606 GAGCTAAATGATCTTATGTTTATACACTGCATTCGTAGTACACACACTGATCAAGG 1665
 QY 250 GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer 263
 Db 1666 AATATGTCGGMACTCTTAAAGACAAATTTTCCGCGGATGGGAATATATCAAGCTCA 1725
 QY 264 ValGlnSerArgAspGlnLeuArgGlyAlaPheValLeuGlyMetLysAspLysGlu 283
 Db 1726 ATTAAAGCCTTGAACAGATTGCTATCTGATTAACAGCTGGTGGTAAGCTTAATCA 1785
 QY 284 ProLysAlaAlaLeuSerGlu-----GluThrAspTyrPhe---AspAlaTyrLys 299
 Db 1786 GCAAGATTTTATTTAGTGAAGTCTGTGAAGTGGGATACGCTTCTGATGCTCAGAAA 1845
 QY 300 AlaIleLysSerAlaSerTyrSerGly----- 308
 Db 1846 CAAATATCTCAATTCGATAGCTGTAATTAATCAATTAATCCGTTTAAATGCAATGATG 1905

QY 309 -----AlaAlaLeuAsnAlaGlyLys----- 315
 Db 1906 AACAACTTCTCTTGTGCTCAGAAATGCGGCTAAACCTGCCCTTAACCTACAGCAAGTCT 1965
 QY 315 ----- 315
 Db 1966 TGGAGTGAGCAGCAAAAGTATGCAAGATAGTCTCAACAGCTAGGGTAATTAAGCTTCAAAAT 2025
 QY 315 ----- 315
 Db 2026 AACTCAGTAATTTGCTATTCAGCTTTCATGCTTTTATTAAGCAGGATTAATGAA 2085
 QY 316 -----ArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal--- 332
 Db 2086 TTACTACACCGCAGGTTGTTTGGCTTAACGCTTCTACAGGAGTAATCAAAATCAGTTGGG 2145
 QY 333 -----ArgSerLeuValSerAlaThrSer-----LeuThrLysAsnGly 345
 Db 2146 TTGCTACCTCCCTTTTATAGCTGACGTAGACACGCAACCCCTTTTGCCTCAGTAAGAAATACC 2205
 QY 346 LeuAlaLeuAlaGly-----GlyTyrAlaGlyValSerLysLeuGlnLysMet 361
 Db 2206 CGCACATTAACCGACAGCTAATTTTGGGCACACGTGCAATGGGGCAA---GAAACTTTA 2262
 QY 362 AlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValAspGln----- 377
 Db 2263 GCGACTGCTGGCTAGCAAGCTGTGATGACTGTGACAGACTGCGCTCAGACAGTTTAAA 2322
 QY 378 -----LeuSerAsnLeuValGlySerValGlyValPheAlaGly 390
 Db 2323 ACTGCTTTCGAGGGTTGCTTGTTCATCTTATGTTGGCGTGATTTGGCTGTGGGA 2382
 QY 391 TyrThrThrAlaGlyLeuAlaThrAspProAla---ValLysLysAla----- 405
 Db 2383 TGGGCGCTAGAAATCATTAATTTCTTCTTTCGAGAACTTAAAGCTTAAGATGATTT 2442
 QY 406 -----GluSer 407
 Db 2443 GAGCAGACGCCAGCAAAACCAATGTCGAAGCAATTAACCAATTAAGCTCCTCAGTATAA 2502
 QY 408 PheIleGln-----AspLysValLysSerThrAlaSerSerThrThrSer 422
 Db 2503 CTATATACAGAAATATAAGAGCTTCACAAAAGTTAAAGAGTCAGATCTTAACTTCAGAT 2562
 QY 423 TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val 435
 Db 2563 GAAGAGCAAAATAATCTTCAAGTCACTCACCAATTAACCAAACTTCCCTGATTAAGT 2622
 QY 436 Lys-----AspMetSerGlyLysAlaIleSerSerThrGlyAlaSerLeuAspSerThr 453
 Db 2623 AAAGGCTATGATTCYCAAGGAATGCAATTCCTTAAGACAAATAAAGAGCTTGAAGAAAGCG 2682
 QY 454 ValAsnAsnLeuArg 458
 Db 2683 ATTGAGAAATCTAA 2697
 RESULT 3
 US-08-194-290-6
 ; Sequence 6, Application US/08194290
 ; Patent No. 5500353
 ; GENERAL INFORMATION:
 ; APPLICANT: Smith, John
 ; APPLICANT: Bingle, Wade H
 ; TITLE OF INVENTION: Bacterial surface protein expression
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shlesinger, Arkwright & Garvey
 ; STREET: 3000 South Eads Street
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: USA
 ; ZIP: 22202
 ; COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/194.290
APPLICATION NUMBER: US/08/194.290
FILING DATE: 09-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Garvey, George A
REGISTRATION NUMBER: 17737
REFERENCE/DOCKET NUMBER: 5946-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-684-5600
TELEFAX: 703-836-5288
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3300 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Caulobacter crescentus
STRAIN: CB 15
US-08-194-290-6

Alignment Scores:
Pred. No.: 0.000602 Length: 3300
Score: 132.00 Matches: 101
Percent Similarity: 36.51% Conservative: 75
Best Local Similarity: 20.95% Mismatches: 166
Query Match: 5.48% Indels: 140
DB: 1 Gaps: 23
US-09-825-414-66 (1-487) x US-08-194-290-6 (1-3300)
QY 22 SerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArgAsp 41
DB 224 TCGAGCGCGCTGCGCTGACCAACCTGAGAGCTGCACAGCAGCGGCTGTGGC 283
QY 42 LeuGlnAlaIleThrAspTyrLeuYshHisValPheAlaAlaHisArgPheSerVal 61
DB 284 ATCCAGACC-----TACCAAGTCTTCACGGCGCTGCCCGCTGGGCC 325
QY 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
DB 326 GCTGG-----CTGACCTTCTG 343
QY 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluGlyGluThr 95
DB 344 GTGCACTCGACCAACCAACGACGCTGACGAGCGCTACTACGCAAG----- 394
QY 96 ProAlaThrIleAlaGluThrPheAlaAlaGlyLysPhe-----AspArg 111
DB 395 -----TTCCGCTCAGAAACCGCTTCACTCACTTCGATCAAC 433
QY 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla----- 128
DB 434 CTGGCAGCGGGCGCGCGCGCG-----GGCAGCGCTTTCGCGCGCGCTACAGCGGC 487
QY 129 ----SerValLeuGlnTyrMetGlnProAlaIleAsnLys-----GlyAspTyrLeuAla 145
DB 488 GTTTCGTACGCCACAGCGCTCGCACCGCTATGACAGATCATCGCAACGCGCGCG 547
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
DB 548 ACC-----GCCCGTGGCGTGCAGCTC 568
QY 166 ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
DB 569 GCGGCGCGCGCTTCTCTGAGCGCGCGCGCAACATGACGACTACCTGACC----- 619

QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGluArg 205
DB 620 -----GCCCTCGGCGCGCAACAGCGCGCTTCAACGCGCGCT 655
QY 205 GlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArgThr 225
DB 656 GCCGACATGACATGCTGGCCCTCAAGGGCGCGCTGATCGGCACCATCTGAAACCGCGCACG 715
QY 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
DB 716 GTGTGCGGCGATCGGTGTGATACCGACCGCGCGCGCGCGCATGATCAACACACCTGTGCGAC 775
QY 243 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
DB 776 GGGCGCGCTTCGACCGCAACAGCGGCGTGGCGAACCCTGTCAACCGCTTACCGCTGCG 835
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSer---ArgAspGlnLeuArgGly----- 272
DB 836 GCGGTGTGCGGTTCGACCTTCGTGACACCGCGCACCGACACCTGACGGGCGCACCGCGC 895
QY 273 -----GlyAlaPheValLeuGlyMetLysAsp 281
DB 896 AACACAGACAGCTTCGCTCGGGGTGAAGTCGCGCGCGCTGCCACCTGACCGTGGCGAC 955
QY 282 LysGluProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaTyrLysAlaIle 301
DB 956 ACCCTGACGCGGCGTCTGCGACCGACGCTGCAAGCTGGGTGCAAGCT----- 1003
QY 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 321
DB 1004 -----CCTGCGGTTCAGCGCT-----CTGCGG 1024
QY 322 LeuAspValAlaThrAspGlyLeuLysValAlaValArgSerLeuValSerAlaThrSerLeu 341
DB 1025 ACCGCGTACAGATCTCGGGCATTCGAACAGATG---AACGTACGTCGGCGCGCTGCGATC 1081
QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeu----- 358
DB 1082 ACCCTGAAAC-----ACGTCTTCGGCGGTGACGCGGTGACGCGCGCTGAAACCAAC 1132
QY 359 -----GlnLysMetAlaThrLysAsn 365
DB 1133 ACCAGCGCGCGCTCAACCGCTCACCGCGCGCGCTGCCAACAACCTGACCGCCACGACC 1192
QY 366 IleThrAspSerAlaThrTyrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
DB 1193 GCGCGTCAAGCGCGCAACAGCTCGCGCTGACGAGCGCGCGCGCACAGCTACCGCTGCTG 1252
QY 385 ValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
DB 1253 ACGGCGGTGACCTCGGGCAGCACGACGCTGCGC-----GCCAACTCGCGC 1297
QY 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
DB 1298 GCTTCGGGCAACGCTGCGGTGAGCGTGCAGCACTCGACGACCAACGCGCGCGCTATC 1357
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyAlaIleSer 444
DB 1358 GCC---GTGACCGGTGTGACGGCGTGAACGCGGTCAACAGCGCGCAACCGCGTGAAC 1414
QY 445 SerThr 446
DB 1415 ACCACG 1420

RESULT 4
US-08-614-377A-6
Sequence 6, Application US/08614377A
Patent No. 5976864
GENERAL INFORMATION:
APPLICANT: Smit, John
APPLICANT: Bingle, Wade H.
APPLICANT: No. 5976864e1ini, John F.
TITLE OF INVENTION: EXPRESSION AND SECRETION OF

```

: TITLE OF INVENTION: HETEROLOGOUS
: TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson PC
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02110-2804
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/614,377A
: FILING DATE: 12-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/194,290
: FILING DATE: 09-FEB-1994
: CLASSIFICATION: 435
: APPLICATION NUMBER: US 07/895,367
: FILING DATE: 09-JUNE-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Tsao, Y. Rocky
: REGISTRATION NUMBER: 34053
: REFERENCE/DOCKET NUMBER: 08106/002001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-542-5070
: TELEFAX: 617-542-8906
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3300 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Caulobacter crescentus
: STRAIN: CB 15
:
: US-08-614-377A-6
:
: Alignment Scores:
: Pred. No.: 0.000602 Length: 3300
: Score: 132.00 Matches: 101
: Percent Similarity: 36.51% Conservative: 75
: Best Local Similarity: 20.95% Mismatches: 166
: Query Match: 5.48% Indels: 140
: DB: 2 Gaps: 23
:
: US-09-825-414-66 (1-487) x US-08-614-377A-6 (1-3300)
:
: Oy 22 SerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerGlyArgAsp 41
: Db 224 TCGAGCGCGCGCTGCGCTGCAACACCTGAGCTGCTAACAGACGACGCGCTGTGCC 283
: Oy 42 LeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisArgPheSerVal 61
: Db 284 ATCCAGACC-----TACCAAGTCTTCACCGCGGCTGCCCGCCGTCGCGC 325
: Oy 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
: Db 326 GCTGCT-----CTGCACTTCTCG 343
: Oy 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluIuGluThr 95
: Db 344 GTGCACTGACCAACCAACCAACACCACTGAAACGACGCGTACTACTCGAAG----- 394
: Oy 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe-----AspArg 111

```

```

: Db 395 -----TTGCGTCAGAAACCGCTCATCACTTTCATCAAC 433
: Oy 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAla----- 128
: Db 434 CTGGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 487
: Oy 129 ---SerValLeuGlnTyrMetGlnProAlaIleAsnLys-----GlyAspThrLeuAla 145
: Db 488 GTTTCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
: Oy 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
: Db 548 ACC-----GCCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 568
: Oy 166 ValGluThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
: Db 569 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 619
: Oy 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGluArg 205
: Db 620 -----GCCCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 655
: Oy 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
: Db 656 GCGCAATGATCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 715
: Oy 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
: Db 716 GTGTGGGCGATCGGTGTGTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 775
: Oy 243 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
: Db 776 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 835
: Oy 256 GlyPheGlyAspArgMetLeuSerValGlnSer---ArgAspGlnLeuArgGly----- 272
: Db 836 GCGGTGTGCGGTTCGACCCCTTCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
: Oy 273 -----GlyAlaPheValLeuGluMetLysAsp 281
: Db 896 AACAAAGACACGTGCGTTCGCGGTGAGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955
: Oy 282 LysGluProLysAlaAlaLeuSerGluGluThrAspThrLeuAspAlaThrLysAlaIle 301
: Db 956 ACCGTGACGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1003
: Oy 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGluPro 321
: Db 1004 -----GCTGCGGTTCAGCGCT-----CTGCGCG 1024
: Oy 322 LeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
: Db 1025 ACCGCGTCAAGTCTCGCGCATCGAACAAGT---AACGTACGTCGCGCGCGCGCGCGTAC 1081
: Oy 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerTyrLys----- 358
: Db 1082 ACCCTGAAC-----ACGTCTTCGCGCGTACGCGGTCTACCGCGCTGAAACCAACAA 1132
: Oy 359 -----GlnLysMetAlaThrLysAsn 365
: Db 1133 ACCACGCGCGCGCGCTCAACCGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1192
: Oy 366 IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
: Db 1193 GCGCGTCAAGCGCGCAACACGTGCGCGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1252
: Oy 385 ValGlyValPheAlaGlyThrThrThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
: Db 1253 ACGGCGTACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1297
: Oy 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
: Db 1298 GCTTGGGCGACGCGTGTGCGTGAAGCGTGGCAACTGAGACAGACACCAACGCGGCGGTATC 1357

```

[illegible]

OY	146	ThrProLeuLysProLeuThrProLeuIseSerGlyAlaLeuSerGlyAlaMetAspGln	165
Db	548	ACC-----	568
OY	166	ValGlyThrLysMetMetAspArgAlaArgLysAspleuHisTyrlSerTherPro	185
Db	569	GCGGGCCGCCGTTCCTTCTAGCGCCGACCAACAATGCATACGTACGACC-----	619
OY	186	AspLysLeuHISAspAlaMetAlaValSerValLysArgHisProAlaLeuGlyArg	205
Db	620	-----GCTTCCTGGCGCCCAACAGCGGTTCCACGGCCGCT	655
OY	206	GlnValValaspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArgThr	225
Db	656	GCCGACATCGACTGCGCCGCTCAAGCGCCGCTGATCGGCACCATCTCGAACGCCGCACG	715
OY	226	ValLeu-----AlaProAlaLeuAlaSerArgrProserValGlnGlyAlaValAsp	242
Db	716	GGTGCGGCGATCGGTGATTACGGACGCCACCGCGCGGAGATGATCAAGACCTGTGGAC	775
OY	243	PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla	255
Db	776	GCGCGCCTGTCGACCGACAGACGCGGCTGCGCTGAACCTGTTCACCGGCTATCCGCTGTCG	835
OY	256	GlyPheGlyAspArgMetLeuSerValGlnSer---ArgaspGlnLeuArgGly-----	272
Db	836	GCGGTGTGGGTTCGACCCCTCTGCTGCACACCGGACCGCACCTTCAGGGACCGCC	895
OY	273	-----GlyAlaPheValLeuGlyMetLysAsp	281
Db	896	AACAACGACAGCTTGCTGGCGGGTAGTAAGTCCCGCGCTCGACCGTACCGTTGGCGAC	955
OY	282	LysGluProLysAlaAlaLeuSerGluGlnThrAspTrpLeuAspAlaTyrlLysAlaIle	301
Db	956	ACCTGACAGCGGGGCTGCGACCGCAGCTCCTCAACTGGGTGACAACT-----	1000
OY	302	LysSerAlaSerTyrlSerGlyAlaAlaLeuAsnIlaGlyAsnGlyMetAlaGlyLeuPro	321
Db	1004	-----GCTCGGGTTCGCGT-----CTCCCG	1021
OY	322	LeuAspValAlaThrAspGlyLeuLysAlaValArgSerIleuValSerAlaThrSerLeu	341
Db	1029	ACCGCGCGACGATCTCGGGCATCGAACAAGT---AACGTGACGTGGGGCGCTGCATC	1089
OY	342	ThrLysAsnGlyLeuAlaLeuAlaGlyIleTyrlAlaGlyAlaSerLysLeu-----	356
Db	1089	ACCTGCAGC-----ACGTCCTCGGGCTGACGGGTCTGACGCGCTGAACACCAAC	1131
OY	359	-----GlnLysMetAlaThrLysAsn	365
Db	1133	ACCACGGCGCGGCTCAAACCGTCACCGCGCGGCTGCGCGAACCCTGACCGCCACGAC	119
OY	366	IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer	384
Db	1199	GCGCGTCAAGCGCGCAACAACGTGCGCTCACGGCGCGCCAACTGACCGCTGCTCG	1253
OY	388	ValGlyValPheAlaGlyTrpThrAlaGlyLeuAlaThrAspProAlaValLysLys	404
Db	1253	ACGGCGGTGACTCGCGGACGACGACCGGTGGC-----GCCAATCTGGCC	1292
OY	406	AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrlVal	422
Db	1299	GCTTGGGCGACCGTGTGCGTAGGCTCGCGAATCGAGACAGACACACCGGCGCTATC	1353
OY	423	AlaAspGlnThrValLysLeuAlaAlaLysThrValLysAspMetSerGlyAlaAlaIleSer	444
Db	1358	GCC---GTACCCGGGTGTACCGCCGTGTACCTGTGCTCAAAAGCGCGGCAACGCCGTATAC	1411
OY	445	SerThr	446
Db	1415	ACCACG	1420

```

RESULT 6
US-08-591-079-9
: Sequence 9, Application US/08591079
: Patent No. 5972899
: GENERAL INFORMATION:
: APPLICANT: Zychlinsky, Arturo
: APPLICANT: Chen, Yajing
: TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORRISON & FOERSTER
: STREET: 2000 Pennsylvania Avenue, NW
: CITY: Washington
: STATE: DC
: COUNTRY: USA
: ZIP: 20006-1812
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/591,079
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Llywyt, Shmuel
: REGISTRATION NUMBER: 33,949
: REFERENCE/DOCKET NUMBER: 15661-20017.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 887-1500
: TELEFAX: (202) 887-0764
: TELEX: 90-4030 MRSNROBRSMSH
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5393 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Salmonella typhi
: STRAIN: Ty2
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 543..2324
: OTHER INFORMATION: /gene="slpP"
US-08-591-079-9

Alignment Scores:
Pred. No.: 0.00152 Length: 5393
Score: 131.50 Matches: 112
Percent Similarity: 36.24% Conservative: 79
Best Local Similarity: 21.25% Mismatches: 167
Query Match: 5.46% Indels: 169
DB: 2 Gaps: 26

US-09-825-414-66 (1-487) x US-08-591-079-9 (1-5393)
QY 3 ILeasngInSerAlaGInGInProPogIyAlAlaMeGluSerPhearGthralaSer 22
| | | | | : : : | | | | | : : : | | | | |
DB 2373 ATAAATCCCGCGCGATTAATATATCATCTGTGGAATAGATTCACAGACAGCTTG 2432
| | | | | : : : | | | | | : : : | | | | |
QY 23 AspAlaSerLeuAlaSerSerSerValArgSerValSerThThSerGysArg----- 40
| | | | | : : : | | | | | : : : | | | | |
DB 2433 CAATCGGTAGCGCTAAAGATATCTGAATAGTATTGATTAGCAGCAGTAAGTCACT 2492
| | | | | : : : | | | | | : : : | | | | |
QY 41 AspLeu----- 42
| | | | | : : : | | | | | : : : | | | | |
DB 2493 GACCTGGGTGAGTCTACTAGCAGCGCGCTGCCGACGGGTATTAAACGAAACCCC 2552
| | | | | : : : | | | | | : : : | | | | |
QY 43 GlnAlaIleThrAspTyrLeuIleYshIshIshValPheAlaIaIaIshArgPheSerValIle 62
| | | | | : : : | | | | | : : : | | | | |

```

```

DB 2553 GGAACGATCAGCTCTTTTAAAA-----GCCAGTATTCAA 2588
QY 63 GlySerProAspGluArgAspAla-----AlaLeuAlaHisAsnGluInIleAspAlaLeu 81
| | | | | : : : | | | | | : : : | | | | |
DB 2589 AATACCGACATGATGATGATTTGAATGCCCTGGCAAAATAT----- 2630
| | | | | : : : | | | | | : : : | | | | |
QY 82 ValGluThrArgAlaAsnArgLeuTyrSerGluGluThrProAlaThrIleAlaGlu 101
| | | | | : : : | | | | | : : : | | | | |
DB 2631 GTACGACTAAAGCGAATGAGTT-----GTCCAAACCCAGTTACCGCGAG 2675
| | | | | : : : | | | | | : : : | | | | |
QY 102 ThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThralaSerSerAlaPheGlu 121
| | | | | : : : | | | | | : : : | | | | |
DB 2676 CACGACGACAGAAGTCGGAAGTTTGTGATATAGCGAATGCTTCCAGTGGCGTT----- 2732
| | | | | : : : | | | | | : : : | | | | |
QY 122 AsnThrProPheAlaAlaAlaSerValLeuGluTyrMetGlnProAlaIleAsnLysGly 141
| | | | | : : : | | | | | : : : | | | | |
DB 2733 ---GCCGTGTGGCTCCCGAATACGTAATGCTG-----ACGTTGAACCAAGCT 2780
| | | | | : : : | | | | | : : : | | | | |
QY 142 AspThrPheAlaThrProLeuLysProLeuThrProLeuIleSer----- 156
| | | | | : : : | | | | | : : : | | | | |
DB 2781 GAT-----AGCAAACTGTCTGTGAAGTTGCTATGCTGATTGATGACAGCTAAA 2831
| | | | | : : : | | | | | : : : | | | | |
QY 157 -----GlyAlaLeuSerGlyAlaMetAspGln 165
| | | | | : : : | | | | | : : : | | | | |
DB 2832 ACGACGGCAAGCTCCATGATGCCGGAAGGATGAATGCGTGTGCGGTGATTTCCAG 2891
| | | | | : : : | | | | | : : : | | | | |
QY 166 -----ValGlyThrLysMet-----MetAsp 172
| | | | | : : : | | | | | : : : | | | | |
DB 2892 AGCGCCCTTCAGTTGGGATCATCTGCGCGGCCCAAACTGGAAATATAGGGGCTGCAG 2951
| | | | | : : : | | | | | : : : | | | | |
QY 173 ArgAlaArgGlyAspLeuHisTyrLeuSerThSerProAspLysLeu----- 188
| | | | | : : : | | | | | : : : | | | | |
DB 2952 AATGAAGAAGGCGCGCTTAAACATAATGCCGACAGATGATGATACCTGACACTGAAC 3011
| | | | | : : : | | | | | : : : | | | | |
QY 189 HisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGluArgGlnAlaVal 208
| | | | | : : : | | | | | : : : | | | | |
DB 3012 CACAGTATTAAACGTCGTGAACGGCGCAGAAATAGCGTCAAACTGTGCTGTAAGCGTC 3071
| | | | | : : : | | | | | : : : | | | | |
QY 209 AspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaIleArgThrValLeuAla 228
| | | | | : : : | | | | | : : : | | | | |
DB 3072 GAT-----TCTCGAAATCGTTAAATATGAAGAAAC----- 3104
| | | | | : : : | | | | | : : : | | | | |
QY 229 ProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyValSerThrAla 248
| | | | | : : : | | | | | : : : | | | | |
DB 3105 -----GTRACCGATGCGACGAAATACTTAATGATGCGACCTTAATATTAAT 3152
| | | | | : : : | | | | | : : : | | | | |
QY 249 GlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAsp 268
| | | | | : : : | | | | | : : : | | | | |
DB 3153 GCCGGAACCAAGCGCCACGGAAGCTGTGGTATTAA-----AACGTAATTAA 3200
| | | | | : : : | | | | | : : : | | | | |
QY 269 GlnLeuArgGlyGlyAlaPheValIleuGlyMetLysAspLysGluProLysAlaAlaLeu 288
| | | | | : : : | | | | | : : : | | | | |
DB 3201 CAATC-----TCCCTGTAACATCAGCTATTCTG 3230
| | | | | : : : | | | | | : : : | | | | |
QY 289 SerGluGluThrAspThrPheAspAlaTyrLysAlaIleLysSerAlaSerLysGly 308
| | | | | : : : | | | | | : : : | | | | |
DB 3231 TCGAAACGCTTACGCTGTGCAATCCGATATTCGCTTGACGAGAAATACCATGATAG 3290
| | | | | : : : | | | | | : : : | | | | |
QY 309 AlaAlaLeuAsnAlaGlyLys---ArgMetAlaGlyLeuProLeuAspValAlaThrAsp 327
| | | | | : : : | | | | | : : : | | | | |
DB 3291 ACCCGAATGAGTCGCGGCAAGATGCAGATGACGCGC----- 3326
| | | | | : : : | | | | | : : : | | | | |
QY 328 GlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla 347
| | | | | : : : | | | | | : : : | | | | |
DB 3327 -----GATGTGATTATGAAGACTCATGCTACG----- 3353
| | | | | : : : | | | | | : : : | | | | |
QY 348 LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThr 367
| | | | | : : : | | | | | : : : | | | | |
DB 3354 ---GTGCGTGTATTGACAGGCGGTCCAG---CAATACCGCCGCTACAG----- 3398
| | | | | : : : | | | | | : : : | | | | |
QY 368 AspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyVal 387
| | | | | : : : | | | | | : : : | | | | |
DB 3399 ---GACGTTCCGACGACCAAAATTACCGAGTGAATAAACCGGTTGCCAGC----- 3446
| | | | | : : : | | | | | : : : | | | | |

```

QY 388 PhealaglytrpThrThrAlaGlyLeuAlaThrAspProAlaValIysAlaGlySer 407
Db 3447 -----ACCGCATCGGACGAGACCCGTGAAGTCACTGTAATTCAGCAGC 3491
QY 408 PheIleGlnAspIysValIysSerThrAlaSerSerThrThrSerTyValAlaAspIle 427
Db 3492 CTGATTCAGGAATGCTCAAAACATGAGAGC----- 3524
QY 428 ThrValIysLeuAlaIysThrValIysAspMetSerGlyIuAlaIleSerSerThrGly 447
Db 3525 -----ATTACCAAGTCGAAAGCATCGCAGCTGCTGCTTCGCA 3563
QY 448 AlaSerLeuArg-----SerThrValAsnAsnLeuArgHisArgSerAlaProGluAla 465
Db 3564 GGCATATTCGGCTTATCTGACAGATCACTATACGCCAT----- 3605
QY 466 AspIleGluGlyGlyIle 472
Db 3606 -----CAGGGGGGAT 3617

RESULT 7
US-09-453-702B-72
; Sequence 72, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
; NUMBER OF SEQUENCES: 265
; CORRESPONDENCE ADDRESS:
; ADDRESS: Quarles & Brady
; STREET: 1 South Plunkett Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B
; FILING DATE: 03-Dec-1999
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955
; FILING DATE: 04-Dec-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296, 95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46819
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-453-702B-72

Alignment Scores:
Pred. No.: 0.0765 Length: 46819
Score: 130.00 Matches: 111
Percent Similarity: 35.89% Conservative: 62

Best Local Similarity: 23.03% Mismatches: 232
Query Match: 4 Indels: 78
DB: 4 Gaps: 18

US-09-825-414-66 (1-487) x US-09-453-702B-72 (1-46819)

QY 5 GlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSerAspAla 24
Db 32171 CAGTCAGCGCTTCCAGCGCAGCAAGCATCAACAAAGGCTACTGAAAGTCAAAAGT 32230
QY 25 SerLeuAlaSerSerSerValArgSerValSerThrThrThrSerGlyArg----- 40
Db 32231 GCTGCCGCTGCAAGTCTCTCAAAAGCGCGCGGCTTACAG-TGCCGCTGCGGCAAAAC 32289
QY 41 -----AspLeuGlnAlaIleThrAspTyrIleuIysHisIysValPheAlaHis 57
Db 32290 GTCGAAAGCAATGGCGAGGTGTCCACAAATTCAGCCGCAC-----TTCGATCCAC 32343
QY 58 Arg-----PheSerVal-IleGlySerProAspIuArgAspAlaAlaHisAs 75
Db 32344 CGCGACACAGAAAGCGTCAGAAAGCTCCTCCAGCCAGGATGCGTGGCT-----TC 32397
QY 75 ngIuGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluTh 95
Db 32398 AAAAGAGCGGCCAAATCATCAGAAACGAGCGACCTCGAGCGCCATGTGACACCTC 32457
QY 95 rProAlaThrIleAlaGluThrPheAlaIysAlaGluIysPheAspArgLeu---AlaTh 114
Db 32458 CTCGGCAACGGCGGCGAGCATTCGCGCAAGCGCCCAAAACGTCGTGAGCAAAACGCTA 32517
QY 114 rThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMe 134
Db 32518 GTCTCTGAAACGCGAGAGAAACAGAGTGCCTCCGACACAGCAGGC----- 32563
QY 134 tGlnProAlaIleAsnIysGlyAspTyrPheAlaIleThrProIleuysProLeuThrProle 154
Db 32564 -----TCAAAACAGCGGCTGCATT 32583
QY 154 uIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrIysMetAspArgAl 174
Db 32584 ATCTGCCAGTGGCGGTCATCAACAGTGGCGGCGGCGCTCAGCCACT----- 32629
QY 174 aArgGlyAspLeuHisTyrIleSerThrSerProAspIysLeuHisAspAlaMetAlaVal 194
Db 32630 -----GCCACCGCGCGGAAATTCGCGAGAAATGCCGATC 32667
QY 194 lSerValIysArgHisSerProAlaLeuGlyArgGlnValIysAspMetGlyIleAlaVal 214
Db 32668 GTCTGCTTCAACAGCCACAGCAAGGCTGGCGAAGCCACTGAACAGCGCAGCAGCAGC 32727
QY 214 lGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerAr 234
Db 32728 GAGTTCTGCTTCGCGAGGAGACATCCGAAACGAAAGCGTTGGAACACAGCGC 32787
QY 234 gProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyIleValAlaAs 254
Db 32788 AGAATCTCAAAAGGCTGCGCATGCTCAAGCCAGTTCGGCGGCTCATCGCATCATC 32847
QY 254 nAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgGspGlnLeuAlaGlyGlyAl 274
Db 32848 TGGCTCT-----GCTTCAAAAGATGAGCGGACGACAGCAAGC 32883
QY 274 aPheValIleuGlyMetIysAspIysGluProIysAlaAlaLeuSerGluGluThrAspTyr 294
Db 32884 GTCAGCAGCAAGAGACAGCGCAGCGCATCCAGAAAGCGACAGAGGAGCGAGCTGAT 32943
QY 294 pLeuAspAlaTyrIysAlaIleIysSerAlaSerTyrSerGlyAla-----AlaLeuAs 312
Db 32944 TGCAGCGCAGCAGCTCAGACCAAAAGTACGGGAATCTCAGACAAACGCGCGTGAGAC 33003
QY 312 nAlaGlyIysArg-----MetAlaGlyLeuProIleuAspValAlaThrAspGl 328
Db 33004 AGCGGCAAAAGCGGCGAGAGATTCATTCGCGCGCTTGAGATGCGAGCAGCAGC 33063

[illegible]

RESULT 8

```

US-09-103-840A-2/c
: Sequence 2, Application US/09103840A
: Patent No. 6294328
:
: GENERAL INFORMATION:
: APPLICANT: FLEISCHMAN, Robert D.
: APPLICANT: WHITE, Owen R.
: APPLICANT: FRASER, Claire M.
: APPLICANT: VENTER, John C.
:
: TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
:
: FILE REFERENCE: 24366-20007.00
: CURRENT APPLICATION NUMBER: US/09/103,840A
: CURRENT FILING DATE: 1998-06-24
:
: NUMBER OF SEQ. ID NOS.: 2
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 2
:
: LENGTH: 4403765
:
: TYPE: DNA
:
: ORGANISM: Mycobacterium tuberculosis
:
: FEATURE:
:
: OTHER INFORMATION: CDC 1551
:
: OTHER INFORMATION: "n" bases at various positions throughout the sequence
:
: US-09-103-840A-2

```

Alignment Scores:

Pred. No.:	244	Length:	4403765
Score:	127.50	Matches:	125
Percent Similarity:	38.64%	Conservative:	86
Best local Similarity:	22.89%	Mismatches:	212
Query Match:	5.30%	Indels:	124
DB:	4	Gaps:	27

US-09-825-414-66 (1-487) x US-09-103-840A-2 (1-4403765)

QY 8 GlnGlnProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAla 27

Db 1261064 GAGCACCCACCGGTGCGAGCTGCGAGTCCGAAGACTGCCAGCCGTGCCGCGACGCC 1261005

QY 28 Ser-----SerSerValArgSerValSerThrThrSerCysArg----- 40

Db 1261004 GCGCCGGCGCTCATCGCCTCGTGGAATGGCGTTTGCCCGTGCCCTGTTGCCGGTGCA 1260945

40 ----- 40

Db 1260944 TCGGTGTGCACCTCAATCGGATCGGCTGCGCGTCCGCGGGCGGTGAACCAACCGTCGATC 1

41 ASPLEUGINALLETHTRASPTYRLLEULSHSHSVAlpheAlaAlaHISArgpHeSer 60

Db 1260884 GAAGGACAGCACGTGACCCAGCCTGTACGTCTCAACCTTACCGCA-----ACC 1

QY 61 valIleGlySerProaspGluargaspAlaAlaLeuAlaHisAsnGluGlnIleaspAla 80

Db 1260833 ATCACCGGCTCCCG-----CGCATCGGCCCGCGCGGAACTC----- 1

QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAlaThrIleAla 10

Db 1260794 -----AAGCGCGCCACCGAAGGCTACTGGGCCGACGTACCAAGCCGATCCGAGCTG 1

QY 101 GLuThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 12

Db 1260743 GAGGCCGTCGCCGCCACGTTACGCCCGCGACACCTGGTGGGCCCTGGCCGCGCGGCTCG 1

Qy 121 GluAsnThrProPheAlaAlaSerValLeu---GlnTyrMetGlnProAlaIleasn 13

Db 1260683 GACTCGGTGCCGGTGAACACCTTCTCCTACTACGACCAATGCTCGATACCGCGGTGCTG 1

QY 140 LysGlyAspTrrPLeuAlaThrProLeuLysProLeuThrProLeuIleSerGly----- 15

Db 1260623 CTCGGC-----GCGCTGCCGCGCCCGGA--GTGAGCCCGGTTCCGACGGGCTGGAC I

QY 158 -----AlaLeuSerGlyAlaMetaspInValGly-----ThrLysMet 17

Db 1260575 CGTATTTCGCCGGCGGGCCACCGACGATCGCGCCGCTGGAGATGACGAAGTGG 1

QY 171 MetaspargalaargGlyaspLeuHISTyrLeu-----SerThrSerPro 18

Db 1260515 TTCGACACC-----AACTACCACTACTGCTGTACCCGAGATCGGGCCGTCGACCACG 1

QY 186 ASPLYSLEUHTSASPALAMETALavalSerValLysArgHISserProAlaLeuGIlyarg 20

Db 1260464 TTCCACGCTGCACCCCGGCAAGGTGCTCGCCGAACCTCAAGAG-----GCCGTTAGGGCAA T

206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 22

Db 1260410 GGCATTCCCGCAGTCCGGTGATCATCGGGCCGATCACCCTTCCTGCTGCTGAGCAAGGCC 1

226 ValLeuAlaProAlaLeuAlaSerArg--ProSerValGlnGlyAlaValAspPheGly 24

Db 1260350 GTCGACGGCGCGGGGC-GCCGATCGAACGCCCTCGAAGAGTTGGTTCGGTCTATTCGA

245 ValSerThrAlaGly-----GlyLeuValAlaAsnAlaGlyPhe 250

DD 1260291 GCTGCTGTCGCTTGGCCGACGGCGGGCCCGCCAGTGGGTGCAGTTCGACGAGCCGGCGT

258 G1YasparqmetLeuserValGInserargspGInLeuargG1YGLYAlaPheValLeu 2/

D5 1260231 GGTGACCGACCTCTCCCCGACGGCCCGCCCTGGCTGAAGCGT-----1

27 2/8 GlymetlysaspIysGluProlYsAlAlAlaLeuSerGluGlu-TirAspTirPleuAspAl 23

DD 1260186 -----GTACACCGCGTGTGCTCGGTGAGCAACCGG-----1

291 arylLysalaaleLysSerAlaSerIyr-----SerGIYAlaAlaLeuashAlaGL 311

DB 1260135 ---CCTGCGATCTATGTCGGCCACCTCATCTTCGGGACCCGGGGCCCTACCCGGG--1

```

314 YLYSARGMETALAGLYLEUPROLEUasp---ValAlaI'hraspGLYLeuLysAlaValAr 33
      ::::| |:::| |

```

DB 1260101 -----CTGGCTCGCACCCCGTCTGAAGCCATCGGCGTCGAC-----1

333 gserleuvalseralainrserleuinrllysasniglyleu---AlAlleuAlaglygly-- 35

Db 1260065 ----CTGTGGCCGCTCCGACACCTCGTGGCCGGGTACCCGAGCTGGCCGCAAGAC 1260010
QY 352 ----TyrAlaGlyAlaSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAl 370
Db 1260009 GCTGGTGGCCGGGCTC-----GTGACGGCGGCAACGTGGCCGACCGA 1259965
QY 370 aThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaG1 390
Db 1259964 CTTGGAGCGCGCTTGGGCGACGTTGGCGACCTCTGCTGGCTTGGCGGCTACCGTGGCCGT 1259905
QY 390 yTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlySerPheAlaG1 410
Db 1259904 CTCGACCTCGTGCACACACGACGACGCTACTCGCTGGAACCGGAACGACGCTGA 1259845
QY 410 nasPylsValLysSerThrAlaSerSerThrThrSerYrValAlaAspGlnThrVally 430
Db 1259844 TGACGCGCTTGGCGGAGCTGGCTGGTCCGCGAAAGGTGGC---GAAGTCTGCT 1259788
QY 430 sleuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLe 450
Db 1259787 TCTCGCGCGTCCCTGCTGGCGAC---GGACACGACGCGGTCCGACGAGATCGCTGCTC 1259731
QY 450 uArgSerThrVal-----AsnAsnLeuArgHisArgSe 461
Db 1259730 CCGCGCGCATCGCTCCGCAAGCGCAGACCCCGGTTACAAATGGGCAAAATCCGGGC 1259671
QY 461 rAlaProGluAlaAspIleGluGluGlyGly-IleSerAlaPheSerArgSerGluThrPr 481
Db 1259670 GCGCATCGAGCGCATCTCGCGTGGAGCCACCGCGCAATCCGCCCGACGCGCGC 1259611
QY 481 ropheGlnLeuArg 485
Db 1259610 CAGCCAAAGACGCGC 1259597

RESULT 9
US-08-961-527-71
; Sequence 71, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid

STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-71

Alignment Scores:
Pred. No: 0.101 Length: 32768
Score: 126.50 Matches: 97
Percent Similarity: 36.27% Conservative: 76
Best Local Similarity: 20.34% Mismatches: 249
Query Match: 5.26% Indels: 56
Gaps: 9

US-09-825-414-66 (1-487) x US-08-961-527-71 (1-32768)

QY 14 AlameGlySerPheArgThrAlaSerAspAlaSerSerValArgSer 33
Db 1508 GCTCTGAATCAGCATCAACAGATGCTCGGCTTACGACAGACAGTGGCTTCA 1567
QY 34 ValSerThr-----ThrSerCysArgAspLeuGlnAlaIleThr 46
Db 1568 GCAAGTACTAGTCATCGGCTTCAACATCGACACAGATGGCTGATCGCATCAAGAGT 1627
QY 47 AspYrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAsp 66
Db 1628 GCTTCGCTCAGCATCAACGAGTGTGACGCTCAGCAGACACATCAGTCTTGAATCT 1687
QY 67 GluArgAspAlaAlaLeuAlaHis---AsnGluGlnIleAspAlaLeuValGluThrArg 85
Db 1688 GCATCAACAGAGTCCGCTTCAACGCTCAACCACTGGCTGGCTCAGCTCGCAAGT 1747
QY 86 AlaAsnArgLeuYrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLys 105
Db 1748 GCTTCGCTTCAAGCATCAACGAGTGTGCGGCTCAGCAAGGCAAGTACCTCAGCAGTCA 1807
QY 106 AlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProhe 125
Db 1808 GCTTCC-----GCCTCACACAGTGTGCTGCTTCAAGCAACAGCATGTCG 1852
QY 126 AlaAlaAlaSerValLeuGlnLysMetGlnProAlaIleAsnLysGlyAspYrLeuAla 145
Db 1853 TCAGCTTACGACAGATATCTCAACGCTTGAATCGCATCAAGAGGCTTCGACGACGA 1912
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db 1913 TCA-----ACGAGTACGTACAGCTTCAACGACATCAGCTTGAATCT 1957
QY 166 ValGlyThrLysMetMetAspArgLarGlyAspLeuHisTyLeuSerThrSer-Pr 185
Db 1958 GCATCAACAGTGTGCTCAGCTCAGCA-----TCGACAAAGCGCC 1996
QY 185 cAsp-----LysLeuHisAspAlaMetAl 193
Db 1997 TCAGCTTACGACAGATCAAGTGTCTTCAACGCTTCAAGGCTGACAGATGCGGCTCAAC 2056
QY 193 aValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMetGlyIleAl 213
Db 2057 AGTCATCTGAATCGCATCAACGAGTGTGCTCAGGCTCAGCAAGTACT-AGTGCATCAGC 2115
QY 213 aValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSe 233
Db 2116 TTCACATCAACAGATGATCGATGGCTTCAAGCATCAACAGTGTGCTTCAAGGCTCAAC 2175
QY 223 rArgProSerValGlnGlyAlaValaAspPheGlyValaSerThrAlaGlyGlyLeuValAl 253
Db 2176 CAGTGCCTCAGCTTACGACAGATGATGCTTCAAGTGTGCTTCAAGTGTGCTTCAAGC 2235
QY 253 aAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyG1 273
Db 2236 CTCACATCG-----ACAAGTGCCTCGGCTTC 2262
QY 273 yAlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAs 293
Db 2283 AGCAAGACATCAAGCATCTGATCAAGCTCAACAGTGTGCTTCAAGCAAGTACAG 2322

```
QY 293 pTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsn1 313
DB 2323 TCCTTCAGCTTACAGATCAACACAGCGCCCTCGGCTTCAGAACACCTTCGATTC 2382
QY 313 aGlyLysArgMetAlaGlyLeuProLeuAspVal---AlaThrAspGlyLeuLysAlaVala 332
DB 2383 GGCTTCACACACAGCGCTCGGCTTCAGAACACCTTCGATTCGATCGGCTCAACAG 2442
QY 332 LArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTyr 352
DB 2443 CGCCCTCAGCTTCAGATCAACAGAGTCTTCGCTTCAGAACACAGCGCCCTCGGCTTC 2502
QY 352 rAlaGlyValSerLysLeuGlnLysMetAlaThr-----LysAsn1IleThrAs 368
DB 2503 AACATCAACAGATAGCTACGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAG 2562
QY 368 pSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPh 388
DB 2563 TCCTTCAGCTTCAGATCAACAGAGTCTTCAGCTTCAGCTTCAGCTTCAGCTTCAG 2622
QY 388 eAlaGlyTrpThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPh 408
DB 2623 AACATCAACAGATAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAG 2682
QY 408 eIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnTh 428
DB 2683 TCCTTCAGCTTCAGATCAACAGCGCCCTCAGCTTCAGAACACCTTCGATTCGATTCG 2742
QY 428 rValLysLeuAlaLysThrValLysAspMetSerGlyAlaAlaIleSerSerThrGlyAl 448
DB 2743 AACCTTCAGAACAGTGGCTCGGCTTCACCAACAGTGCATCTGAATCGGCTCAACAG 2802
QY 448 aSerLeuArgSerThr-ValAsnAsnLeuArgHisArgSerAlaPro 463
DB 2803 AACCTTCAGAACAGTGCATCTGAATCGGCTTCACCAACAGTGCCT 2849

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.: 314 Length: 4411529
Score: 126.50 Matches: 125
Percent Similarity: 38.648 Conservative: 86
Best Local Similarity: 22.898 Mismatches: 212
Query Match: 5.268 Indels: 124
DB: 4 Gaps: 27

US-09-825-414-66 (1-487) x US-09-103-840A-1 (1-4411529)
QY 8 GlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAla 27
DB 1261536 GAGACACCAACCGGCTGCGAGCTCGAGTCCGAAAGACCTGCACAGCCGCGGAGCGCC 1261477
```

```
QY 28 Ser-----SerSerValArgSerValSerThrThrSerCysArg----- 40
DB 1261476 GCGCCCGCGGCTCATTCCTTCGTGGAAATGGCGCTTGGCCGCTTCGCTTCGCGGTGCA 1261417
QY 40 ----- 40
DB 1261416 TCGGTTCACACTTCATATCGCATTCGGCTCGCGGCTCGCGGCGGTGAACACCGCTGCATC 1261357
QY 41 AspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisArgPheSer 60
DB 1261356 GAAGACACGACGTGACCCACCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAG 1261306
QY 61 ValIleGlySerProAspGlnArgAspAlaAlaLeuAlaHisAsnGlnIleLeuAspAla 80
DB 1261305 ATCACCGGCTTCGCGC-----GCGATCGCGCGCGCGCGCGGAACTC----- 1261267
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGlnGlyGluThrProAlaThrIleAla 100
DB 1261266 -----AAGCGCGCCACGGAAGCTAATGGCGCGGAGCTACCAAGCCGATTCGAGCTG 1261216
QY 101 GluThrPheAlaLysAlaLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
DB 1261215 GAGCCCGTTCGCGCCACGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAGCTTCAG 1261156
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeu---GlnTyrMetGlnProAlaIleAsn 139
DB 1261155 GACTCGGTGGCGGTGAACACCTTCCTCTACTACGACCAATGCTGATACCGCGGTGCTG 1261096
QY 140 LysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGly----- 157
DB 1261095 CTCGGC-----GCGCTGCGCGCGCGCA---GTGAGC-----GCTTCGAGCGGCTGGAC 1261048
QY 158 -----AlaLeuSerGlyAlaMetAspGlnValGly-----ThrLysMet 170
DB 1261047 CGCTAATTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260988
QY 171 MetAspArgAlaArgGlyAspLeuHisTyrLeu-----SerThrSerPro 185
DB 1260987 TTCGACAC-----AATCACTACCTACTGATCGGACCGGAGATCGGCGCGCGCGCG 1260937
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuArg 205
DB 1260936 TTCACGCTGCACCCCGGAGAGGTGCTCGCGCAATCAAAAG-----GCGTTAGGGCA 1260883
QY 206 GlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArgThr 225
DB 1260882 GGCATTCCCGCACGCTCCGATCATTCGCGCGGATCACTTCGCTGCTGAGCAAGGCC 1260823
QY 226 ValLeuAlaProAlaLeuAlaSerArg---ProSerValGlnGlyValAlaAspPheGly 244
DB 1260822 GTCCAGCGCGCGCGGGGCGC-GCCGATGGAACGCTCGAAGAGTTGGTCCGCTTATTCGA 1260764
QY 245 ValSerThrAlaGly-----GlyLeuValAlaAsnAlaGlyPhe 257
DB 1260763 GCTGCTGTGCGTGTTCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1260704
QY 258 GlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyValAlaPheValLeu 277
DB 1260703 GGTGACGCGACCTTCCTCCCGGCGCGCGCGCGCGCGCGCTGCTGAAACGGT----- 1260659
QY 278 GlyMetLysAspLysGluProLysAlaAlaLeuSerGlnGlu-ThrAspTrpLeuAspAl 297
DB 1260658 -----GTACACCGCGCGCTGTGCTGCTGAGCAACCG----- 1260628
QY 297 aTyrLysAlaIleLysSerAlaSerTyr-----SerGlyAlaAlaLeuAsnAlaGly 314
DB 1260627 -----CTTCGATCTATGTGCGCACCTTACTTGGGGAGACCGCGCGCGCTTACCGGCGC-- 1260574
QY 314 YLysArgMetAlaGlyLeuProLeuAsp---ValaIleThrAspGlyLeuLysAlaValaVal 333
DB 1260573 -----CTGGCTGCGACACCGCGGTGCGAAGCATGCGGCGCTGCAC----- 1260538
```

```

QY      333 gserleuvalseralathrserleuthrlylsamcglyleu--alaleualaglygly-- 351
          |||||
Db 1260337 ----CTGGTGGCCGGTGGCCGACACACCTGGTGGCCGGGGTACCAGCCGCGGCAAGAC 12604822
          |||||
QY      352 ----TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAl 370
          |||||
Db 1260481 GCTGGTGGCCGGGTCT-----GTCCAGCGGGCGCAACTGTGGCCACCCA 12604377
          |||||
QY      370 aThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaG1 390
          ::|||::|
Db 1260436 CCTGGAGGCGGGCGGTGGGACAGGTGGGACCTCTCTGGTGGCGGGCGCTACCGTGGCCGT 12603777
          ::|||::|
QY      390 yThrThrThrAlaGlyLeuAlaAlaThrAspProAlaValLysLysAlaGlnuserPheIleG1 410
          |||||
Db 1260376 CTCACAGCTCGTGTCTGCACACTGCACAGTGGCCGTACTCGCTGGACCCGGAACCCAGACCTGGGA 1260317
          |||||
QY      410 nAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLy 430
          TGAGCGGTTGGCGGAGACGTGGCTGGCGTTCCGGTCCGGAAGAAGTGGCC--GAGTCTGTCT 1260260
          |||||
Db 1260316 TGACGCGGTGGCGGAGACGTGGCTGGCGTTCCGGTCCGGAAGAAGTGGCC--GAGTCTGTCT 1260260
          |||||
QY      430 sIeuAlaLysThrValLysAspMetSerGlyGlnAlaIleSerSerThrGlyAlaSerLe 450
          |||||
Db 1260259 TCTGCGGGGTGCTCCCTGGCGCAC--GGACACAGACGCGGTGCGCAGACAGATCGGTGTCT 1260203
          |||||
QY      450 uArgSerThrVal-----AsnAsnLeuArgHisArgSe 461
          |||||
Db 1260202 CCGCGCGCCGCACTCCGTCGCCGCAAGCGCGACCCCGCGGTACACAAATGAGCAATCCGGGGC 1260143
          |||||
QY      461 rAlaProGlnAlaAspIleGlnGlyGly-IleSerAlaPheSerArgSerGlnThrP 481
          |||||
Db 1260142 GCGCATCGAGCGCATGCTGCGCTGCCGAGCCACCAGCGGCAATGCCGCCGCGCGCGC 1260083
          |||||
QY      481 rOPheGlnLeuArg 485
          |||||
Db 1260082 CAGCCAGACGCGC 1260069
          |||||

RESULT 11
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (ERO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; ASnAsnLeuArgHisArgSe
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

Alignment Scores:
Pred. No.: 0.00357 Length: 3666
Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
DB: 2 Gaps: 25

QY      3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaIaMetGlu----- 16

```

Dd	2023	GTAACCAAAAGCGGCAAAAAAGTAACTGGTACTTTATTAAGAAAACAATTAAACAATC	2082
Oy	17	-----Ser	17
Dd	2083	TACAACTACTGGGTANTGATTTAAAGTAGTAACCAAGTTATGTCACCAAAATGCGTACG	2142
Oy	18	PheAtgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThr	37
Dd	2143	TACACAGTAACCTTGAAGAGTACTTTATCTTCA-----GGAAGCTGTAATTACA	2193
Oy	38	SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisIleValPheAlaIleHis	57
Dd	2194	CCTGTAGAAGATTTCAGAAAGTACTTCAAGTCAGATGGTAAACAAACATGCTTTAAAGTAAT	2253
Oy	58	ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGln	77
Dd	2254	GCTACAGATAATTCGTTAATTAACAGCGGTAAGAACTATGCATTTATGCTGTAAGAACST	2313
Oy	78	IleAspAlaLeuValGluThr-----	84
Dd	2314	ACAGCTACATTCAACGCTTACCAAAATGAAGTTCCAAACTCTTACACTGGTGTAACCTCA	2373
Oy	85	-----ArgAlaAsnArgLeuTyrSerGluGlyGluThrPro	96
Dd	2374	TTCAATACACGCTGATTCGTGGTTCACAAACAGCACTATTTGGTGGTTCGGTAAACCCA	2433
Oy	97	AlaThrIleAlaGluThrPheAlaIleValAlaGluLysPheAspArgLeuAlaThrThrAla	116
Dd	2434	GTCAAATATGCTGGTGTATATGACGCCAAACATTAATAATAC-----TTGGAGACTAATGCT	2487
Oy	117	SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr-----	133
Dd	2488	AATGACGATATTGGTGAAGCGCATGGAAGACATTA--TTAACTCAATATGCAACTGAA	2544
Oy	134	-----MetGln	135
Dd	2545	GGCCAAAAAGTAAACAATCTCATTAATGTATGATGTGATATACACTTTAAAGTAATT	2604
Oy	136	ProAlaIleAsnLysGlyAspTrpLeuAlaIleThrProLeuLysProLeuThrProLeu---	154
Dd	2605	AGTGTGTTAATTCCTTCACTGAAGCATATCAAACACAGTTGCTGCCAACAAACACACACACT	2664
Oy	135	--IleSerGlyAlaLeuSer-----GlyAlaMetAsp-	164
Dd	2665	CCAACTACTGGCGCATTAACATTAACACSCAGCAGCTGGTGGTGTGATGATTAAACAAT	2724
Oy	165	GlnValGlyThrLysMetLeuAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer	184
Dd	2725	GCAACTAACACTTTAGCAATTTCTATTAGCTGATGCGATGATCTTAATGTAAAGTGCACAACT	2764
Oy	185	ProAspLysLeuHisAspAlaMetAlaValSerValLys--ArgHisSerProAlaLeu	203
Dd	2785	GTT-----GATACTGCGACAGCGTTCTCATTAAGATAGAGCAAAATTAATCATTA	2832
Oy	204	GlyArgGlnValValaLaspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValVal	223
Dd	2833	TCTCTTAACATTAGTTGAAGCTGGCTAATATACAGGTATTTGCTACA-----	2880
Oy	224	ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValaLaspHe	243
Dd	2881	---ACTGTTCAAGCTGGTACATTATGCTCT------	2907
Oy	244	GlyValSerThrAlaGlyLeuValAlaAsn-----AlaGly	256
Dd	2908	-----TTAATCTGCTGCTAATTAACAAGTTACTTATGACGATGCTAAAAATGCTGCAGCT	2961
Oy	257	PheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly-	273
Dd	2962	GTTGCTGAATAATTAATCTGCTACGCTTAACGTTTAAGAAACAATCTAGGACAAATTAATCTT	3021
Oy	274	--AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThr	292

Db 3022 GATACATTTCACAGGCTGATTA-----CCATCAGACGCTACACGACGTGAATAT 3072
Qy 293 AAPTPLLeuAspAlaTyrLysAlaIleuLysSerLysSerGlyAlaIleu---- 311
Db 3073 ACTTCAATTCATTCCTCAGATTATACATTTCGACACAGGGAAGGATTCACCTTTAAAT 3132
Qy 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
Db 3133 ATTGATTAATGCTGTCGCTCAGATAATTACTTACGA-----GGTAAA 3174
Qy 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
Db 3175 AAGGTGCAACAGGTGCTGCTATGCTATGCTACATGCTACATTTGCACTTGCACAGCTGTT 3234
Qy 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp---- 368
Db 3235 TCTGCA-----GACAAAGTAGTATTATTAAATCACCCTACACAGCT 3273
Qy 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal---- 382
Db 3274 GTTGGTTCGAGTTGAAGTTACATTCTCTCTGTTAATCAGATATTAAATGCAAGTAT 3333
Qy 383 ---GlySerValGlyValPheAlaGlyTyr---ThrThrAlaGlyLeuAlaThrAspPro 400
Db 3334 AAGGTAAAGATCAAGCTGCTGACAGACAGCTGCTACAAAGACATTCACGATTACTACA 3393
Qy 401 AlaValLysValAlaGlySerPheIleGlnAspLysValLysSerThr----- 416
Db 3394 GCCCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3453
Qy 417 -----AlaSerSerThrThrSerTyr 423
Db 3454 GGAAGCTCTCAACAGCAATATCATTCGTAGTTGATCTGCTGCTGCTGCTGCTGCTGCTGCT 3513
Qy 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
Db 3514 GTAGCTGACCAAGCTCAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3573
Qy 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
Db 3574 TTTCACGCTTCTGCAACAGGCTGCTACTATTACATTACTTACTTACTGTA 3621

RESULT 12
US-08-682-517-14
: Sequence 14, Application US/08682517
: Patent No. 5874267
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: Expression of surface layer proteins
: NUMBER OF SEQUENCES: 25
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/682,517
: FILING DATE:
: CLASSIFICATION:
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3666 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3666
: US-08-682-517-14

Alignment Scores: 0.00357 Length: 3666
Pctd. No.:

Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
DB: 2 Gaps: 25

US-09-825-414-66 (1-487) x US-08-682-517-14 (1-3666)

Qy 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGlu----- 16
Db 2023 GTAAACCAAGCGGCAAAAAGTAAGTAACTGGTACTTCTATTAGAAACCAATTATACAAATC 2082
Qy 17 -----Ser 17
Db 2083 TACAACTACGTGCTATGATATTAAAGTAGATTAACCAAGTTATCTACCAAAATGCTAGC 2142
Qy 18 PheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThr 37
Db 2143 TACACAGTAAGTATGAGCTACTTATCTTCTACCA-----GGAAGCTGTATTACCA 2193
Qy 38 SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHis 57
Db 2194 CCTGCTAAAGAAATTTAGAAAGTTACTTCACTGATGATGATGATGATGATGATGATGATGAT 2253
Qy 58 ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGln 77
Db 2254 GCTACAGGTATTCCTCTTATTATACAGACGGTAAGACATATGATCTCTGCTAAAGAAAGCT 2313
Qy 78 IleAspAlaLeuValGluThr----- 84
Db 2314 ACAGCTACATTCACAGCTCAAAATGAAGTCCAAACTCTTACACGTGCTGATGCTACAA 2373
Qy 85 -----ArgAlaAsnArgLeuTyrSerIleGlyGluThrPro 96
Db 2374 TTCATATACAGCTGATTCGTGCTGCTCAACACCACTGATTTGTTGCTGCTGCTGCTGCTGCT 2433
Qy 97 AlaThrIleAlaGluThrPheAlaLysAlaGlyLysPheAspArgLeuAlaThrThrAla 116
Db 2434 GTCAAAATATGCTGCTGCTATACAGCAACCAATATATATATAC-----TTCGAGCTAATGGT 2487
Qy 117 SerSerAlaPheGlnAsnThrProPheAlaAlaAlaSerValLeuGlnTyr----- 133
Db 2488 AATGAAGTATTTGCTGACGCGCATGGAACATTA---TTACTCAATATGCAACTGAA 2544
Qy 134 -----MetGln 135
Db 2545 GGCCAAAAGTACAACTCATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2604
Qy 136 ProAlaIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeu--- 154
Db 2605 AGTCTGTTAATTCCTTCAACTGACATCAACCAACGTTGCTTCAACCAACACGACAGCT 2664
Qy 155 ---IleSerGlyAlaLeuSer-----GlyAlaMetAsp----- 164
Db 2665 CCAACTACAGCGCATTAACATTAACACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2724
Qy 165 GlnValGlyThrLysMetIleAspArgAlaArgGlyAspLeuHisIstyrLeuSerThrSer 184
Db 2725 GCAACTTACACCTTGAAGATTTCATTATGCTGATGACGATGATGATGATGATGATGATGATGAT 2784
Qy 185 ProAspLysLeuHisAspAlaMetAlaValSerValLys---ArgHisSerProAlaLeu 203
Db 2785 GTT-----CATACTGCAACTGTTCTTAAAGAGATGTCCAATATATTCATTA 2832
Qy 204 GlyArgGlnValValAspMetGlyIleAlaValAlaGlnThrPheSerAlaLeuAsnVal 223
Db 2833 TCTCTTACATTAGTGAACCTGGTCTAATACAGGTGATTTGCTACCA----- 2880
Qy 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnIleValAlaAspPhe 243
Db 2881 ---ACTGTTCAAGCTGTACATTATCTCT----- 2907
Qy 244 GlyValSerThrAlaGlyLeuValAlaAsn-----AlaGly 256

```
Db 2908 -----TTAAGTGTGCTACATTAACAGTTACTTATGACAGATGCTAAAAATGCTCAGGT 2961
Oy 257 PheGlyAspArgMetLeuSerValInSerArgAspGlnLeuArgGlyCly----- 273
Db 2962 GTTGCTGAAAAATATTACTGCTAGCGTAACTTAAGAAACACTGACGAACTTACTTCT 3021
Oy 274 ---AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluLutThr 292
Db 3022 GATACATTATTCACAGGCGTATTA-----CCATCAGCAGCGTACAGCAGCTCAAAAT 3072
Oy 293 AspTrpLeuAspAlaTrpLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
Db 3073 ACTTCAATCAATTCGTCAGATTTATACATTTCGACACAGGTGAGGATTCCTTTAAAT 3132
Oy 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
Db 3133 ATTGATTAATGCTGGTGCCTCAAGTAACTTAATTAACCTTAGCA-----GGTAAA 3174
Oy 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
Db 3175 AAAGGTGACACAGGTGATGCTGATGCTATCATGCTACATTGTCAGAGTACGTCACACTGTT 3234
Oy 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
Db 3235 TCTGGA-----GACAAAGTAGTTATTAAATCAACCTCAACAGGT 3273
Oy 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
Db 3274 GTTGCTGCTGAAGTGAAGTTACATCTCTCTGTTAACTCAAGATTAATTAATCAGTAGTT 3333
Oy 383 ---GlySerValGlyValPheAlaGlyTrp---ThrThrAlaGlyLeuAlaThrAspPro 400
Db 3334 AACGCTAAAGATCAAGTGGTTCAGACAGACAGCTGCTCAAAAGCATTCCAGATTACTACA 3393
Oy 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThr----- 416
Db 3394 GCCCTTCTGTGGGTGAAAAAGTAGTTATTGATGGTGGTAATATACCTGCTAGCATTT 3453
Oy 417 -----AlaSerSerThrSerThr----- 423
Db 3454 GGAATGCTCCACACAGCAATTCATTCGATGTAATCGCTGCTCAATATCATTAAGCTTCA 3513
Oy 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyLut----- 441
Db 3514 GAGCTGACAGCTGCAAAATCTTCTGCTACAAATTAATTAATTAACCTGACAGTAAG 3573
Oy 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
Db 3574 TTACAGCTTCTGCAACAGGTGCTACTATTAACATTAACTTACTGTA 3621

RESULT 13
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-7

Alignment Scores:
Pred. No.: 0.00446 Length: 4197
Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
                Gaps: 25

US-09-825-414-66 (1-487) x US-08-682-517-7 (1-4197)
Oy 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGlu----- 16
Db 2207 GTAAACCAAGCGGCAAAAAAGTACTGCTACTTGTATTAAGAAAGCACTTATACATC 2266
Oy 17 -----Ser 17
Db 2267 TACAACTAGTGTGCTAATGATTTAAAGTACATAACCAAGTTATCTCACCAAATCGTAGC 2326
Oy 18 PheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThr 37
Db 2327 TACACAGTAACTTATGAAGCTACTTATCTCTTACA-----GGAACGTATTATACA 2377
Oy 38 SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaIleHis 57
Db 2378 CCGTGAAGAAATTTAGAAAGTTACTTACAGTGGATGCTAAACACACGCTGTTAAAGTATT 2437
Oy 58 ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGln 77
Db 2438 GCTACAGGATATGCTGTTAAATCAAGACGTAAGACATGATTCATTCGCTAAAGAACCT 2497
Oy 78 IleAspAlaLeuValGluThr----- 84
Db 2499 ACAGCTAATTCACAGCTACCAAAATGAAGTCCAAACCTTTCACCTGGTGAGCTAGCTGAA 2557
Oy 85 -----ArgAlaAsnArgLeuTyrSerGluGlyGluThrPro 96
Db 2558 TTCATTAACAGCTGATTCGTGTTCAACACAGCAACCTGATTTGGTTCGCGTAAACCA 2617
Oy 97 AlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAla 116
Db 2618 GTGAATATGCTGCTGATACAGCAAAACATATAAATAC-----TTCGAGCTAATGCT 2671
Oy 117 SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlyLut----- 133
Db 2672 AATGAAGTATTTGGTGAAGCGGATGGGAAGCATTA---TTAACTCAATATGCAACGTGAA 2728
Oy 134 -----MetGln 135
Db 2729 GGCCAAAAGTACAAATGCTATATAATGTAATGCTGATACAGTTACATTAAAGTAAATT 2788
Oy 136 ProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeu--- 154
Db 2789 AGTGGCTGAATTTCTTCAACTGAACGTATCAAAACAGTTGCTCCAAACACACACAGCT 2848
Oy 155 ---IleSerGlyAlaLeuSer-----GlyAlaMetAsp----- 164
Db 2849 CCAACTACTGGGCGACATTAACATTAACACCAGACAGCTGGTGTTAGTTGATTTAAACAACT 2908
Oy 165 GlnValGlyThrLysMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
Db 2909 GCAATTAACACTTTAGAAATTCATTAAGTGCAGATCTTAAGTAAAGTGAAGTGAACAACT 2968
Oy 185 ProAspLysLeuHisAspAlaMetAlaValSerValLys---ArgHisSerProAlaLeu 203
Db 2969 GTT-----GATACGTCGACAGCTGTTTCATTAATAAGATAGGCAATAATTCATTA 3016
Oy 204 GlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValVal 223
Db 3017 TCTCTTACATTAAGTGAAGACGCTGCTAATGACAGGTGATTTGCTACA----- 3064
```

```

OY 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPhe 243
      ||||| ||||| ||||| |||||
DB 3065 ---ACTGTCACAGCTGGTACTATATCTCT-----
OY 244 GlyValSerThrAlaGlyGlyLeuValAlaAsn-----AlaGly 256
      ||||| ||||| ||||| |||||
DB 3092 -----TTAACTGCTGCTGACTTAACTTACTTATGACAGTGTCTAAATAATGCTGCAGGT 3145
OY 257 PheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly----- 273
      ||||| ||||| ||||| |||||
DB 3146 GTTCTGTGAATAATATTACTGCTAGCGTAACATTAAAGAAATACTAGCGACATTACTTCT 3205
OY 274 ---AlaPheValLeuGlyMetLeuAspLysGluProLysAlaAlaLeuSerGlnGluThr 292
      ||||| ||||| ||||| |||||
DB 3206 GATCATATTACACAGAGGTGATTTA-----CCATCAGACGCTACAGCAGCTGAATAT 3256
OY 293 AspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
      ||||| ||||| ||||| |||||
DB 3257 ACTTCTAAATCAATTGCTGCAGATTATACATTTCACACAGGTGAAGCATTCCTTAAT 3316
OY 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
      ||||| ||||| ||||| |||||
DB 3317 ATTGATATATGCTGCTGCTCAGCTAATTAATTACCA-----GGTMAA 3358
OY 330 LysAlaValaArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaAla 349
      ||||| ||||| ||||| |||||
DB 3359 AAGAGTGCACAGGTGACTGATGATCAATGCTATTCATTCGACGATTCGACAGCTGT 3418
OY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
      ||||| ||||| ||||| |||||
DB 3419 TCTGGA-----GACAAAGTACTTATTAATTCAGCTACACAGCT 3457
OY 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
      ||||| ||||| ||||| |||||
DB 3458 GTTGTCTCGAAGTTGAAGTACATTCCTCTCTGTTAATCAAGTATTAATAAGCAGTAGT 3517
OY 383 ---GlySerValGlyValPheAlaGlyTrp---ThrThrAlaGlyLeuAlaThrAspPro 400
      ||||| ||||| ||||| |||||
DB 3518 AACGTTAAAGATCAAGTCGTTGACGAGACAGCTGCTACAAAAGCATTCACGATTACTACA 3577
OY 401 AlaValLysLysAlaGlySerPheIleGlnAspLysValLysSerThr----- 416
      ||||| ||||| ||||| |||||
DB 3578 GCCCTTCTGTGGTGAAGAAAGTAGTATTGATGCTGTGAATATACGCTGATGACATTT 3637
OY 417 -----AlaSerSerThrGlyAla-----SerLeuArgSerThrVal 423
      ||||| ||||| ||||| |||||
DB 3638 GGAAGCTGCTCCACAGCAATATACATTGCTAGTGAATCTGCTGAATATACATTGACTTCA 3697
OY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
      ||||| ||||| ||||| |||||
DB 3698 GTACTGACCAAGCTGCAAACTTCTGCTGCTACATTCATTAAACACGCGCATTAAG 3757
OY 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
      ||||| ||||| ||||| |||||
DB 3758 TTTACAGCTTCTGCAACAGGTGCTACTATTCATTACTTACTCTACTGTA 3805

```

```

RESULT 14
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30 (PRO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:

```

```

; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184
; US-08-682-517-8

Alignment Scores:
Pred. No.: 0.00446 Length: 4197
Score: 125.50 Matches: 117
Percent Similarity: 32.998 Conservative: 73
Best Local Similarity: 20.318 Mismatches: 219
Query Match: 5.21% Indels: 167
DB: Caps: 25

US-09-825-414-66 (1-487) x US-08-682-517-8 (1-4197)

OY 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGlu----- 16
      ||||| ||||| ||||| |||||
DB 2207 GTAACCAAGCGGCAAAAGTAGTACTGTTCTATTAGAAAGCAATTATACAAATC 2266
OY 17 -----Ser 17
      ||||| ||||| ||||| |||||
DB 2267 TACAACTAGTGTGCTAATGATTAATTAAGTAGATTAACCAAGTTATCTCCAAATGCTAGC 2326
OY 18 PheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValaArgSerValSerThrThr 37
      ||||| ||||| ||||| |||||
DB 2327 TACACAGTAATTAAGAGCTACTTATCTGTCA-----GGAAGTGTATTACAA 2377
OY 38 SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisLysValPheAlaHis 57
      ||||| ||||| ||||| |||||
DB 2378 CCTGCTAAGAAATTTAAGAGTTACTTCAAGTGGATGTAACAACTGCTGTTAAAGTAAT 2437
OY 58 ArgPheSerValIleGlySerProAspGluThrArgAspAlaIleAlaHisAsnGluGln 77
      ||||| ||||| ||||| |||||
DB 2438 GCTACAGGTATTGCTGTTAATATACAGACGCGTAAGACATATGATTTACTGCTTAAGAGCT 2497
OY 78 IleAspAlaLeuValGluThr----- 84
      ||||| ||||| ||||| |||||
DB 2498 ACAGCTACATTCACAGCTACACAAATGAAGTTCCAACTCTTACACTGCTGATGACTACAA 2557
OY 85 -----ArgAlaAsnArgLeuTyrSerGlnGlyGluThrPro 96
      ||||| ||||| ||||| |||||
DB 2558 TTCAATACAGCTGATTTCTGTTCAACAGCAACTCTATTGTTGCTGTGTTAAACCCCA 2617
OY 97 AlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAla 116
      ||||| ||||| ||||| |||||
DB 2618 GTGAATATATGCTGGTGTATTCAGCAAAACATATTAATATC-----TTGCGAGCTAATGCT 2671
OY 117 SerSerAlaPheGluAsnThrProPheAlaAlaIleAspValLeuGlnTyr----- 133
      ||||| ||||| ||||| |||||
DB 2672 AATGAAGATTTGTGAAGCGCATGGGAAGCATTTA-----TTAACTCAATATGCAACTGAA 2728
OY 134 -----MetGln 135
      ||||| ||||| ||||| |||||
DB 2729 GGCCAAAGATTAACAATCTCATATATATGATAGTGTATACAGTTAAAGTAAT 2788
OY 136 ProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeu--- 154

```


Db 1121 -----CCTGCTGATGTTAGTGTAGAACATTTCATCANGTGA 1159
Qy 166 ValGlyThrLysMetCAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
Db 1160 GTTGGGAAAAAGTATCTGAAGAACTAGTAAAGAGAA-----AGTACTCCTGAA 1207
Qy 186 AspLysLeuHisAsp-----AlaMetAlaValSerValLysArgHisSerProAla 202
Db 1208 GTTAAAGCAGAAAGATTTCACCACTGCTGTAGATGATAGTGGACATTCATCANGTGA 1267
Qy 203 LeuGlyArgGlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnVal 222
Db 1268 GTTGAGAAAAAGTATCTGAAGAACTAGTAAAGGAAAAATACT----- 1309
Qy 223 ValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
Db 1310 ---CCTGAAGTATGAGCAGAAAGATTG---CAACCTGCTGTAGATGTTAGTGTAGAA 1360
Qy 243 PheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeu 262
Db 1361 CATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAAGAACTAGTAAAGAGAAAGTACTCCT 1420
Qy 263 SerValGlnSerArgAsp-----GlnLeuArgGlyGlyAlaPhe 275
Db 1421 GAAGTTAAAGCAGAAAGATTTCGAACCTGCTGTAGATAGTATGACATTCATCAAGT 1480
Qy 276 ValLeuGlyMetLysAspLysGlu---ProLysAlaAlaLeuSerGlnGluThrAspTrp 294
Db 1481 GAAGTTGGAAAAAGTCTGAAACTAGTAAAGAGAAAGTACTCTGAAGTTAAAGCA 1540
Qy 295 LeuAspAlaTyrLysAlaIle-----LysSerAlaSerTyrSerGlyAla 309
Db 1541 GAAGTTTGCACCACTGCTGTAGATGCTAGTGAACAATTCATCAAGTGAAGTTGGAGCA 1600
Qy 310 AlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro-----LeuAsp 323
Db 1601 AAAGTATCTGAAGAACTAGTAAAGAGAAATACTCTGAAGTTAAAGCAGAAAGATTGCCA 1660
Qy 324 ValAlaThrAspGly-----LeuLysAlaValArgSerLeuValSerAla 338
Db 1661 CCTGCTGATGATGATGCTAGTGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAA 1720
Qy 339 ThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyIleTyrAlaGlyValSerLysLeu 358
Db 1721 ACTACTAAAGAGAAATACCTCCTGAAGTTAA-----GCCGACATTTG 1765
Qy 359 GlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla-----ValSer 376
Db 1766 CAACCTGCTGTAGATGATGCTAGTGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCT 1825
Qy 377 GlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrAlaGlyLeu 396
Db 1826 GAAACTAGTAAAGAAAGTACTCTGAAGTTAAAGC-----GAAGATTTCACACT 1879
Qy 397 AlaThrAspProAlaValLysLysAlaGlnSerPheIleGlnAspLysValLysSerThr 416
Db 1880 GCTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1939
Qy 417 Ala-----SerSerThrThrSerTyrValAla---AspGlnThrValLysLeuAlaLys 433
Db 1940 AGTAAGAAGAAAGTACTCCTGAAGTTAAAGGAGAAAGTTGCCAACCTGCTGTAGATGAT 1999
Qy 434 ThrValLysAspMetSer-----GlyGluAlaIleSerSerThrGlyAlaSerLeu 450
Db 2000 AGTGTGAACATTCATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 2050
Qy 451 ArgSerThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGlu 468
Db 2051 -----AAAGAGAAAGTACTCTGAAAGTTAAAGCGGAA 2083

Pred. No.: 6,63e-240 Length: 1464
 Score: 2407.00 Matches: 487
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0

US-09-825-414-66 (1-487) x US-09-825-414-65 (1-1464)

QY 1 MetHisIleAsnGlnSerIleArgInGlnProGlyValAlaMetClnSerPheArgThr 20
 DB 1 ATGCAATACCAATCCCAACACCGCTGGCGTTGCAATGAGAGTTTCGACA 60
 QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerThrSerCysArg 40
 DB 61 GGTTCGAGCGCTCCCTTCCTCGAGTTCTGTGCGGTCTGTGACGACTACCTGTCGCC 120
 QY 41 AspleuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisArgPheSer 60
 DB 121 GATCTACAAGCTATTACCATTTATCGAAACATCACGTGTGCTGCGCACAGGTTTTCG 180
 QY 61 ValIleGlySerProAspGlnArgAspAlaIleuAlaHisAsnGlnIleAspAla 80
 DB 181 GTAATAGGCTCACCGGATAGCGCTGATGCCGCTTTCACACACAGACGATCGATGCG 240
 QY 81 LeuValGlnThrArgAlaAsnArgLeuTyrSerGlnGlyGluThrProAlaThrIleAla 100
 DB 241 TTGGTAGACACAGCGCCACCGCTGTACTCCGAAGGGAGACCCCGCAACCATCGCC 300
 QY 101 GlnThrPheAlaLysAlaLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
 DB 301 GAAACATTCGCCAAGCGGAAAAAGTTGACCGTTTGGCAGACCCGATCAAGTGCTTTT 360
 QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys 140
 DB 361 GAGAACACGCCATTTCCCTGCCCTGCTGCTTCACTACATGACGCTCGATCAACAAG 420
 QY 141 GlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
 DB 421 GCGGATTTGGCTGACAAAGCGCTCAAGCGCTGACCCCGCTCATTTCCGAGCGCTGCG 480
 QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgLysPheLysIle 180
 DB 481 GAGCGCATGACAGGTGGGACCAAAAGATGATGATGCGAGGGGTATCTGCACTTAC 540
 QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
 DB 541 CTGAGCACTTGCCTGGACAAAGTTGATGATGGATGGCGGTATCGGTGAAGGCCACTCG 600
 QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
 DB 601 CCTGGCGCTTGGTCAGACAGTTCTGACATGGGGAATTGACAGTCAGACGTTCTCGCGGCTA 660
 QY 221 AsnValIleArgThrValLeuAlaProAlaLeuAlaSerValProSerValGlnGlyAla 240
 DB 661 AATGTGGTGGCTGACCGTATGGCTCCAGCATAGGCTCCAGACCGCGGTGAGGGTGTCT 720
 QY 241 ValAspPheGlyValSerThrAlaGlyLysLeuValAlaAsnAlaGlyPheGlyAspArg 260
 DB 721 GTTGATTTTGGGCTATCTACGCGGGGTGGCTTGGTTGCCAATGACAGCTTTGGCAGCCG 780
 QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
 DB 781 ATGCTCAGTGTCAATCCGCGCATCACTGCTGGGGGGGCAATGCTACTTGGCATGAAA 840
 QY 281 AspLysGlnProLysAlaAlaLeuSerGlnGluThrAspTyrPleuAlaLysIleVal 300
 DB 841 GATTAAGACCCCAAGCGCGCTTGGTGAAGAACTGATGCTGATCTTCAACAAGCG 900
 QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
 DB 901 ATCAAGTGGCGCACTACTACAGTGGCGGCTCAATGCGGGCAAGCGGATGGCCGCGCTG 960

QY 321 ProLeuAspValAlaIleThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
 DB 961 CCACGTGAGCTGCCGACCCACCGGCTCAAGCGGGTGAAGAGTCTGTGTCGCCACACAG 1020
 QY 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyIleTyrAlaGlyValSerLysLeuGlnLys 360
 DB 1021 CTGACAAAAAATGGCTTGGCTTACCGGCTGCTTACCGCGGGGTAAAGTTGCGACAAA 1080
 QY 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
 DB 1081 ATGGGACGAAAAATATCTCATTTGCGGACACAGGCTGCGTTAGTCAGCTGACACAC 1140
 QY 381 LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
 DB 1141 CTGGTGGGTGGTGGGCTTTTCGAGGCTGGACACCGCTGGACCTGGCGACTGACCT 1200
 QY 401 AlaValLysLysAlaGlnSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
 DB 1201 CGGTTAAAGAAAGCCGAGTCTTTATACAGATTAAGGTGAATCGACCGCATCTAGTACC 1260
 QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
 DB 1261 ACAAGCTATGTTGCCGACACACCGCTCAACTGCGGAAAAAGCTAGAGACATGAGCGGG 1320
 QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
 DB 1321 GAGCGGATCTCCAGCACCGGCTCCAGCTTACGACGATGCTCAATATACCTGGCTATCGC 1380
 QY 461 SerAlaProGlnAlaAspIleGlnGlyIleSerAlaPheSerArgSerThrLys 480
 DB 1381 TCCGCTCCGGAAGCTGATATCGAAGAGGTGGATTTGGCTTTTCGAAAGTAAACA 1440
 QY 481 ProPheGlnLeuArgArgLeu 487
 DB 1441 CCGTTTACAGCTACGCGCTTTG 1461

RESULT 2

US-09-825-414-6

; Sequence 6, Application US/09825414

; Patent No. US20020083489A1

; GENERAL INFORMATION:

; APPLICANT: Collmer, Alan

; APPLICANT: Alfano, James R.

; APPLICANT: Charkowski, Amy O.

; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE

; FILE REFERENCE: 19603/3243

; CURRENT APPLICATION NUMBER: US/09/825,414

; CURRENT FILING DATE: 2001-04-03

; PRIOR APPLICATION NUMBER: 60/194,160

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: 60/224,604

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: 60/249,548

; PRIOR FILING DATE: 2000-11-17

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 6

; LENGTH: 1461

; TYPE: DNA

; ORGANISM: Pseudomonas syringae

; US-09-825-414-6

Alignment Scores:

Pred. No.: 2.67e-184 Length: 1461
 Score: 1869.50 Matches: 381
 Percent Similarity: 85.19% Conservative: 33
 Best Local Similarity: 78.40% Mismatches: 71
 Query Match: 77.67% Indels: 1
 DB: 10 Gaps: 1

US-09-825-414-66 (1-487) x US-09-825-414-6 (1-1461)

```
QY 1 MethisIleasnGlnSerIleaglInProProGlyValAlaMetGluSerPheargThr 20
   |||||::: ||||| 111 ::|||
Db 1 ATGCACATCAACGACGCGCTCCACACACGCGCTGTGACGCGATGATTCGACACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArg 40
   |||||::: ||||| 111 ::|||
Db 61 GCGTCGACGCGCTCTCTTCCCTCCAGCTGTGCGCATCTGTGACGCTCCGATCAGCAACGC 120
QY 41 AspLeuGlnAlaIleThrAspTyrLeuLysHisIleValPheAlaIleHisArgPheSer 60
   |||||::: ||||| 111 ::|||
Db 121 GAGATTAATGCGATTGCGCATCTACGACAGATCAATGTTGCGTGGCATTAACCGCG 180
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisasnGluInIleAspAla 80
   ||| 111 ::| 111 ||||| 111
Db 181 CCGGCGCATTCGCGTCATGCGCCACAGCTGACGTTGACGTACACATCCGACGATCATCGG 240
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
   |||||::: ||||| 111 ::|||
Db 241 CTGATCGACAGCGCGCGCCGCGCTGACATTGCGAGGGGAAACCCCGGCAACCATCGCC 300
QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerAlaPhe 120
   |||||::: ||||| 111 ::|||
Db 301 GACACCTTCGCGCAAGCGGGAAGCTCGACCGATGGCGACGACATCAATCAGCGCGTTG 360
QY 121 GluAsnThrProPheAlaAlaAlaSerValIleGluInTyrMetGluInProAlaIleAsnLys 140
   |||||::: ||||| 111 ::|||
Db 361 CCGGCGACGCCCTTTCGCCATGCGCTGCTGCTTCACTACATGACGCGCTCGATCAACAG 420
QY 141 GlyAspTyrPheAlaIleThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
   |||||::: ||||| 111 ::|||
Db 421 GCGGATTCGCTCGCGCTCGCTCGCTCAAAACGCTGACCCGCTCATTTCCGAGCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
   |||||::: ||||| 111 ::|||
Db 481 GCGCGCATGACGACGCGGCGCACCAAGATGATGGACCGCGCGCATGATCTGCATTTAC 540
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
   |||||::: ||||| 111 ::|||
Db 541 CTGACGCGCTCGCGCGACAGGCTCCCGATGCGATGGCGCTTCGCTGAACGCCACTCG 600
QY 201 ProAlaLeuGlyArgGlnValAlaAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
   |||||::: ||||| 111 ::|||
Db 601 CCAAGCGCTGCTGACAGGTTCTGACACGCGGGGTTGCGGTTCCAGACGATCGCGCGCG 660
QY 221 AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
   |||||::: ||||| 111 ::|||
Db 661 AACGCCGTAACGATATGGCTCGCGCACGCGCTCCAGACCCCGCTGCGCGGTGCT 720
QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
   |||||::: ||||| 111 ::|||
Db 721 GTGGACCTTGCTGTATCGATGCGGGGTGCTGCTGCCAACGCGCTTTGGCAACCGC 780
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
   |||||::: ||||| 111 ::|||
Db 781 CTGCTCAGTGTGACAGTGGGTATCCACAGCGTGGCGGTGATTAATGCTCGGTTGAAG 840
QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaThrLysAla 300
   |||||::: ||||| 111 ::|||
Db 841 GATTAAGACCCCAAGCGCTCACTGACCGCAAGAAACGATGCTCCAGCTTATAAGCA 900
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
   |||||::: ||||| 111 ::|||
Db 901 ATCAAAATGCGCACGCTACTCGGCTCGCGCTCAACGCTGGCAAGCGGATGGCGGCTCG 960
QY 321 ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
   |||||::: ||||| 111 ::|||
Db 961 CCACTGATATGTCGACGACGCAATGGGTGGGTAAAGACCTGTTGTCAGGTCACAC 1020
QY 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyLysSerLysGluInLys 360
   |||||::: ||||| 111 ::|||
Db 1021 CTGACCCCAAAAGGCTGTGGCTCGCGGTGCTTGCAGGGGTAGCAAGTTGCGAGAG 1080
QY 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
```

```
Db 1081 ATGGCAGCAAAAATATCAACCGCCGCGACCAAGCGCGGTGATGATTCGATTCACAC 1140
QY 381 LeuValGlySerValGlyValPheAlaGlyTyrPheThrThrAlaGlyLeuAlaThrAspPro 400
   |||||::: ||||| 111 ::|||
Db 1141 CTGGCAGTTCGCGCACGCCCTTTTCGCGAGCTGTGACCCGCGCGCTGCAACCCGATCCC 1200
QY 401 AlaValLysAlaGlyLeuSerPheIleGlnAspLysValLysSerThrAlaSerThr 420
   |||||::: ||||| 111 ::|||
Db 1201 GCGGTAAAAAAGCCGAGTCTTCATACAGACACGCTAAATCGACTCATCTCAGTACC 1260
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
   |||||::: ||||| 111 ::|||
Db 1261 ACAGGCTAGTAGCGCACGACGCGCAACGCTGCGAAGACCGTCAAAACATGCGCGG 1320
QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
   |||||::: ||||| 111 ::|||
Db 1321 GAGGCGATCAACCATACCGCGCGCCAGCTTGCCTCAATACGCTCAATMACTCGCTCAACGC 1380
QY 461 SerAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArgSerGluThr 480
   |||||::: ||||| 111 ::|||
Db 1381 CCGGCTGTGAAGCTGATATAGAAAGGGGCGACGCGCGCT--TCTCAAGTAAATA 1437
QY 481 ProPheGlnLeuArgArg 486
   |||||::: ||||| 111
Db 1438 CCGTTTCGCGCTATGCGG 1455

RESULT 3
US-09-825-414-1/c
; Sequence 1, Application US/09825414
; Patent No. US20020083489A1
;
GENERAL INFORMATION:
; APPLICANT: Colimer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
;
SEQ ID NO 1
;
LENGTH: 30365
;
TYPE: DNA
;
ORGANISM: Pseudomonas syringae
;
FEATURE:
; NAME/KEY: unsure
; LOCATION: (29734)
; OTHER INFORMATION: n at any position is undefined
US-09-825-414-1

Alignment Scores:
Pred. No.: 1,95e-182 Length: 30365
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
Gaps: 1

US-09-825-414-66 (1-487) x US-09-825-414-1 (1-30365)
QY 1 MethisIleasnGlnSerIleaglInProProGlyValAlaMetGluSerPheargThr 20
   |||||::: ||||| 111 ::|||
Db 23984 ATGCACATCAACGACGCGCTCCACACACGCGCTGTGACGCGATGATTCGACACA 23925
QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArg 40
```

```
Db 23924 GCGTCGACGGGCTCTCTCCCTCCAGCTCTGCGATCTGTCAGTCCGATCAGCAAGC 23865
Oy 41 AspleuglnAlaIleThrAspTyrIleuYshHisValAlpheAlaAlaHisArgPheSer 60
Db 23864 GAGATAAATGCGATTCGCCGATTACTGACAGATCATGTGTTCCTCCGCTCAAACTGCCG 23805
Oy 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
Db 23804 CCGGCGCATTCGGCTGATGGCCAAAGCTGACGTGACATCAATAGCCAGATCACTGCG 23745
Oy 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAlaThrIleAla 100
Db 23744 CTGATCGAGACGGCGCGCAGCGCTGCACTTCGAAAGGGAACCCCGCAACCATCCAC 23685
Oy 101 GluThrPheAlaIleAlaGluAspPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 23684 GACACCTTCGCCCAAGCGGAAAGCTGACCGATTCGGCAGACATCACTACAGCCCGTTG 23625
Oy 121 GluAsnThrProPheAlaAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsn 140
Db 23624 CGGCGCAGCCCTTTCGCTGATGGCTGCTGATGATGACAGCTGCGATCAACAG 23565
Oy 141 GlyAspThrPleuAlaIleThrProLeuYshProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 23564 GCGGATTCGCTGCGCGCTCCGCTCAACCGCTGACCCGCTCATTTCCGAGCGCTGTCG 23505
Oy 161 GlyAlaMetAspGlnValGluTyrThrLysMetLysAspArgAlaArgGlyAspLeuHisTyr 180
Db 23504 GCGCGCATAGACACAGTGGGACCAAGATGATGACCCGCGCAGCGGATGATGATGATGAT 23445
Oy 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 23444 CTGAGCGGCTGCGCGCAGACGCTCCAGATGATGATGATGATGATGATGATGATGATG 23385
Oy 201 ProAlaLeuGluArgGlnValValAlaPheMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 23384 CCAGCCCTTCTGCTCGACAGATTCGAGACGCGGGTGGCTTCACAGCTACTCGCGCGCGC 23325
Oy 221 AsnValAlaArgThrValLeuAlaAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
Db 23324 AACGCGCATGACGATGATGCTGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCT 23265
Oy 241 ValAspPheGlyValSerThrAlaGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
Db 23264 GTGGACCTGTTGATTCATGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 23205
Oy 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
Db 23204 CTGCTCAGTGTGACATGCTCGATCACCAGCGTGGCGGTGATGATGATGATGATGATG 23145
Oy 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspThrPleuAspAlaTyrLysAla 300
Db 23144 GATAAAGACCCCAAGCTCAGCTGAGCAAGCAAGCAAGCAAGCTGCTGAGGCTTATTAACA 23085
Oy 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 23084 ATCAATATGCGCAGCTACTCGGGTTCGCGCTCAACGCTGCGCAACGCGATGGCGGTCGT 23025
Oy 321 ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
Db 23024 CCACGATATGCGCAGCGACGCAATGGGTGGTAAAGCTGCTGACGCTGACGCTGACG 22965
Oy 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyValSerLysLeuGlnLys 360
Db 22964 CTGACCCCAAAAGCTCTGCGCTGCGGCGGTGCTTGCAGGGGTGAGCAAGTGTGAGAG 22905
Oy 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 22904 ATGGGAGCAAAATATACCGACCGCGCAACAGCGCGGTGATGATGATGATGATGATGAT 22845
Oy 381 LeuValGlySerValGlyValAlaPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
```

```
Db 22844 CTGCGAGTTGGGACCGCTTTTCGACAGCTGACACGCGCGCTGACACAGCATCC 22785
Oy 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
Db 22784 GCGGGAAGAAAGCCGAGTCTTCTATACAGACACAGCGGAAATGACATGATCCAGTACC 22725
Oy 421 ThrSerThrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 22724 ACAGGCTACGTAGCCGACCAACGCTCAAACTGCGAAGACCGTCAAAAGACATGGCGCG 22665
Oy 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnArgHisArg 460
Db 22664 GAGGCGATCCCATACCGGCGCGCAGCTTGGCAATACGTCATACCTGATACCTGTCACACG 22605
Oy 461 SerAlaProGluAlaAspIleGluGluGlyLysIleSerAlaPheSerArgSerGluThr 480
Db 22604 CCGGCTGCTGAAGCGATATAGAAAGAGGGGGGCAACGCGGCT---TCTCCAAAGTGAATA 22548
Oy 481 ProPheGlnLeuArgArg 486
Db 22547 CCGTTTCGCGCTATGCGG 22530

RESULT 4
US-09-971-536-27
; Sequence 27, Application US/0971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Hayukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christenson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 4645
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-27

Alignment Scores:
Pred. No.: 2,16e-05 Length: 4645
Score: 147.00 Matches: 104
Percent Similarity: 33.53% Conservative: 64
Best Local Similarity: 20.76% Mismatches: 241
Query Match: 6.11% Indels: 92
DB: Gaps: 12

US-09-825-414-66 (1-487) x US-09-971-536-27 (1-4645)
Oy 1 MethsIleAsnGlnSerAlaGlnProProGluValAlaMetGluSerPheArgThr 20
Db 2213 ATCAAGCGCATATAGTGTGCTGATGTCACAGTTCACAGCAAGGCAAGGCAAGGCT 2272
Oy 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArg 40
Db 2273 GCTAGTGTGCGGCGAGATTATGCAAGTCAAGCAATCATGATTCCTGCGAGTCAATGCTGAT 2232
```

```

Oy 41 AspleuGlnAlaIleThrAspTyrLeuYsHISValPheAlaAlaHisArgPheSer 60
   :::::::::::
Db 2333 AATATGGA-----ATCAAGTCTCTCCGACGATGCTAGAGACCAATCG 2377
Oy 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
   |||
Db 2378 CAATATGCTTTGGCAGTACGACAAAGTCTGCTGGCGCTAGTTCACAGCGGCACGCGCGCA 2437
Oy 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAlaThrIle--- 99
   ::::
Db 2438 GCAAATCGTG-----GCAAGTGGCCGCGCTAGTGAAGCGTCACTGCACTCTCGCGCTA 2491
Oy 100 -----AlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
   |||
Db 2492 AGTAACGCTGATGCATACAGCAAACTCTGCAGCGCGCTTATGATGATCTTACCGCTCTGAG 2551
Oy 116 AlaSerSerAla-----PheGluAsnThrProPheAlaAla 127
   |||
Db 2552 GCCAGTGGCGGCTTGTGCTGATGATAGTTCGGGATATGCGGCATGACATGATTCGACGA 2611
Oy 128 AlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeuAlaThrPro 147
   ::::
Db 2612 AGTTCGCTGCGCGTGCATG----- 2632
Oy 148 LeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 167
   |||
Db 2633 -----AGCCGACGCTTATGCACAGCCCAAGCTCTGCCAAG 2668
Oy 168 ThrLysMetLeuAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLys 187
   ::::
Db 2669 GTTCGAGTGCAGTGCAGCAGCAGCGGGT----- 2698
Oy 188 LeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnVal 207
   |||
Db 2699 -----ACTGCAGCTGCTGTGCTAGTGCAGCTCAAGGCGACTCCAAAGATTAACAAGCG 2752
Oy 208 ValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArg-----Thr 225
   |||
Db 2753 ACTGCAGCTACAGCAAGAAGTCAACGACTTGATGATTTGAATGAATCAAGTCTTAAGT 2812
Oy 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly----- 239
   |||
Db 2813 GATTACGAAAGTGGCGCAAGCTCAGTGCAGCAAGCGGTCAGCAAGCATCGATCGACA 2872
Oy 240 -----AlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsn 254
   |||
Db 2873 TCTGCGTATGCTAGTGCCTGCAAGTTGAGTGCAGTGAAGCGGTCATATATCTCATCAG 2932
Oy 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274
   |||
Db 2933 GCAGGCTCCAGCGGCAGGAGCGCTCGTCCGTCAGTCCGCGAGTGCAGCCCAACATCGCAC 2992
Oy 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294
   |||
Db 2993 ACCGCTGCGAGTGCAGTATCCGAGATAGGATGAGGATTCAGTCACTATAGCCACT 3052
Oy 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
   |||
Db 3053 CAGGCTGAAGGAGGAGCAAGCAAGCAAGCAAGTAAAGCGAGTCCGCAACGCGCCGCGC--- 3109
Oy 315 LysAlaGluAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSer 334
   |||
Db 3110 -----GCGCGCGCTTGGTTCAGTCTGCTGCAGTGCAGTGAAGGCAAGCGCGAAGAG 3160
Oy 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGly 354
   |||
Db 3161 GCTGCAGATGCGATGCTGTCAGACAGTGCAGCGCAGCAGCAAGTAAACAGTAAATGCAAGT 3220
Oy 355 ValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla 374
   ::::
Db 3221 GCCGCAAGC-----AGTCCGCAAGGAGCTGCT 3247

```

```

Oy 375 ValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrAla 394
   |||
Db 3248 GATACGAAGCGCGCA-----GCAGATTCCTCGAGTGA 3280
Oy 395 GlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLys 414
   |||
Db 3281 GCGAGTGTGCAGCAAGCAGTGCACAAAGGTGCGAAGACAGTGTCCGACGAGCGGAGT 3340
Oy 415 SerThrAlaSer-----SerThrThrSerTyrValAlaAspGlnThrValLysLeuAla 432
   ::::
Db 3341 GCCGGCGATCCGATGACCTGCTAGCTTCTAGTCCGCGCAGTCCGCGCTTCAGCTTTCAC 3400
Oy 433 LysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyValaSerLeuArgSer 452
   |||
Db 3401 AAAGCTGCCAGCGCTGCGGAAGCGCAGCTTCAAGTGCAGCGCGCGCTGCTACTTCA 3460
Oy 453 ThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGluGlyIle 472
   |||
Db 3461 GCGGCGATCAGAGGACACAGAGTGGCGGCAAGCTCCAGTGCAGCGAAGCGGTCAGCA 3520
Oy 473 Ser 473
   |||
Db 3521 TCA 3523

RESULT 5
US-09-813-214A-8
; Sequence 8, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Ploisla, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE,
; FIDE REFERENCE: 7969-089-999
; CURRENT FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 08/968,685
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 9542
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-8

Alignment Scores:
Pred. No.: 6.75e-05 Length: 9542
Score: 146.50 Matches: 111
Percent Similarity: 36.00% Conservative: 69
Best Local Similarity: 22.20% Mismatches: 181
Query Match: 6.09% Indels: 139
DB: 9 Gaps: 23

US-09-825-414-66 (1-487) x US-09-813-214A-8 (1-9542)
Oy 32 ArgSerValSerThrThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuYsHIS 51
   ::::
Db 1346 AAGACTTTTAACTATCAGCAGTGTGCACAGACAGCAGCGCTTAACCGAT----- 1393
Oy 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
   |||
Db 1394 -----CATTAACATCGGTGTGTGTAACAATAAGGCGATGTGTGAAGTTCA 1438
Oy 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArg-----AlaAsnArg 88
   |||
Db 1439 CTTGCTGAACCTTAAACAGCTTAATAATGTTACCACTGAATAACCTAACCCCAACGAG 1498
Oy 89 LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGlyLys 108
   ::::
Db 1499 AAAGTTAACGTAGCAAAAC----- 1519
Oy 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla 128

```

Db 1520 -----CGCCTT-----ACCACGAGATAAATGTGTTTACCAGATGAT-----ATG 1558
Qy 129 SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeu 148
Db 1559 AATGGCATGTGTAAGAACCAAACTTATCTGTATTAAGACACTGGCATTCATGACAGTGGT 1618
Qy 149 LysProLeuThrProLeuIleSerGlyValLeu----- 159
Db 1619 CAAAAGATTACCAAACTTACGTGCTGTAGTACATGACGATGGCGCAACTTATGACAG 1678
Qy 160 -----SerGlyAlaMetAspGlnValGlyThrLysMetMet 171
Db 1679 CTTAAAAAGTTAAACCAACCGCTGAAGGTCTGACCAAACTTATACCGTTAAAAAGSTA 1738
Qy 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 1739 GATAAAAATGTATGATGAT-----GCTAATGACAGC 1768
Qy 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMet--- 210
Db 1769 AAATTCATCACCCTGGCGTAAATAATACAAACAGCGTACTCAAGTCAACACCTTAAA 1828
Qy 211 -----GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1829 CTCAAAGGTGAAGGCGTGTGATGTACAAAC----- 1861
Qy 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal 245
Db 1862 -----GAACCAAAATGGTACAGTGTACCTTTGGCGCT 1891
Qy 246 SerThrAlaGlyGlyLeu-----ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerVal 264
Db 1892 AACCAAAATACGCGTCTACCGCTGGCAACAGCACCTTAACCAAGAGGCTTATCTGTT 1951
Qy 265 Gln-----SerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAsp 281
Db 1952 AAAAAACCAATAGTAAACAAACAAATCAAGTGGCTGATGACATTACATTTACTGAT 2011
Qy 282 -----LysGlnProLysAlaIleLeuSerGlnGluThr-----AspTrp 294
Db 2012 ATCAGCAATAGTAAAGCCGCGCTGGCAATGAAATACCACTGCATTAACGCGATTAAGAG 2071
Qy 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 2072 ATT-----GGTTTGGCTAATTAATCTGCT 2095
Qy 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAsp-----GlyLeuLysAla 331
Db 2096 TCATTGGATGCAAAACAAACCCGCTTACCCCACTGCATTAACGCGATTAAGAG 2155
Qy 332 ValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 351
Db 2156 CTGACCAATGTCACATTCGCTTAAACCTCTCACTCACTGCT----- 2197
Qy 352 TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThr 371
Db 2198 ---GGGCGACTAGACTTATATACCGCTTAAAGCACTGCTAATACCGAAAAATCAGGCT 2254
Qy 372 LysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrp 391
Db 2255 GCCCGCACCATTAAGACTTATACCACTTATCACAAGTACAGCGTACCTTGACAGCT--- 2311
Qy 392 ThrThrAlaGlyLeuAlaThrAspProAlaVal---LysLysAlaGluSerPheIleGln 410
Db 2312 -----GATACAGGTCTTAATGTACCCAAAAAATCGCGGAGATTTTAAAG 2356
Qy 411 AspLys---ValLysSerThrAlaSerSerThrThr-----SerTyrValAla 425
Db 2357 GTTAAAGGTGTAAACCAACGCTGATCATTTAAACCAAAATTAACATCGGCTGTGCT 2416
Qy 426 AspGln-----ThrValLysLeuAlaLysThrValLysAspMetSerGly 440

Db 2417 GATAGTACCGATTAATAGCTTAACCGTTAAACCTTAAACCTTAAAGGATCTTGATGCC 2476
Qy 441 GlnAlaIleSerSerThrGlyValSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 2477 GTTATATACCTAAACCTTACCTACGCGCATTAAGTATACCGTAAAGACAGTGGCAACAC--- 2533
Qy 461 SerAlaProGlnAlaAspIleGlnGluGlyIleSerAlaPheSerArgSerGluThr 480
Db 2534 -----ACCGCTAACCTCAAAATGTGATTTAAACC---TTAGCAACCAAAATGACA 2581
RESULT 6
US-09-932-183A-1
Sequence 1, Application US/09932183A
Patent NO. US20020127641A1
GENERAL INFORMATION:
APPLICANT: Estell, David A.
TITLE OF INVENTION: Proteases From Gram-Positive Organisms
FILE REFERENCE: GC394C1-05
CURRENT APPLICATION NUMBER: US/09/932,183A
CURRENT FILING DATE: 2001-08-17
PRIOR APPLICATION NUMBER: US 09/308,375
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US98/18828
PRIOR FILING DATE: 1998-09-08
PRIOR APPLICATION NUMBER: EP9719636.4
PRIOR FILING DATE: 1997-09-15
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 7100
TYPE: DNA
ORGANISM: Bacillus subtilis
US-09-932-183A-1
Alignment Scores:
Pred. No: 0.000209 Length: 7100
Score: 140.00 Matches: 128
Percent Similarity: 33.44% Conservative: 81
Best Local Similarity: 20.48% Mismatches: 222
Query Match: 5.82% Indels: 194
DB: 10 Gaps: 26
US-09-825-414-66 (1-487) x US-09-932-183A-1 (1-7100)
Qy 15 MetGlnSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerVal 34
Db 862 ATGGAATTATATACAGCGACGACGACCAAGTAATGTCAAAACCTTAATACACGGTATGCC 921
Qy 35 SerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisIleValPhe 54
Db 922 AGTCTCTAGGCGCTCAGTAATAGACAAAGCTGTCAAGTATTTTGAAAT----- 969
Qy 55 AlaAlaHisArgPheSerVal---IleGlySerProAspGluArgAspAlaIleAlaLeuAla 73
Db 970 GCAGTAATATAGCTTAAATGTAAAGCAGCTGGAAGCAATATATACAGATGCACAAATTCAAAGC 1029
Qy 74 HisAsnGlnGlnIleAspAlaLeu-----ValGlnThrArgAlaAsnArgLeuThr 90
Db 1030 TTGAATATGCAATTTTGAAGAAATAGCTCCAAACCTCAAAACGCTGAATCAAGCCTCT 1089
Qy 91 SerGlnGlyGluThrProAlaThrIleAlaGluThrPhe----- 103
Db 1090 TCTTTTGA-----GCAGAACTAACCCAAACCTTCAAAACAGCATGTCACCTATTTA 1140
Qy 104 -----AlaLysAla 106
Db 1144 ATCTCCGCTTTTATCTACGAGCTATCTGCACTTAAAGAAATGATATCCAGGCA 1200
Qy 107 GluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla 126
Db 1201 ATGAAATATCTCTCATGACCAAAATATTCGCGGTGTTAAGATGAGCCGGATTTAA 1260
Qy 127 AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThr 146

```
Db 1261 TATATGAACTTCTC-----CAAGATCTATGACTTAAAGTCATACACTTTCAAT 1311
Qy 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
Db 1312 AAAATCAGACAGATTTCTTCAATGACAGCGCATTTTGGAGAAATGGGTTTCGATGAAAGT 1371
Qy 167 GlyThrLysMetAspArgGlyAlaArgGlyAspLeuHisTyrLeuSer---ThrSerPro 185
Db 1372 GAGCTCTCCACGTTAAAGCAAACTGCCCAAGTTCTTCAAAATGTCSTGATTTAACTCC 1431
Qy 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 1432 GATGATACAGTTAAACACGCTTAACGGCAGCAATGCTCAACTTTAAATTTGCAGCAAAATGAT 1491
Qy 206 GlnVal-----ValAspMetGlyLeuAlaValGlnThrPhe 217
Db 1492 TCAATATCAATTCAGCATAAATTAATGAGGTTGATTAATACATATCTGTCTCAACCTCTA 1551
Qy 218 SerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerVal 237
Db 1552 GATCTGGCCCAATTTCTATCCGTA---GCTGGTTCACTGCTTCTACATTTCCGGGGTA 1605
Qy 238 GlnGlyAlaValAspPheGlyValSerThrAla-----Gly 249
Db 1606 GAGCTAAATGATCTTATGTTATACAACTGCATTCGTAACACAGCTGAATCAGGG 1665
Qy 250 GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer 263
Db 1666 AATATCTCGGGAGACCTCTTAAAGCAAAATTTTCGCCGGATTCGGGAATATCAAAAGCTCA 1725
Qy 264 ValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGlu 283
Db 1726 ATTTAAAGCTTTGACACAGATTGGTATCTCAAGTTAAACACGCTGGTGGTGAAGCTTAATGA 1785
Qy 284 ProLysAlaAlaLeuSerGlu-----GluThrAspTrpLeu---AspAlaTyrLys 299
Db 1786 GCAAGGATTTATATGATGAAGTGGCTGGTAAGTGGGATACGCTTCTGTCGTCAGAAA 1845
Qy 300 AlaIleLysSerAlaSerTyrSerGly----- 308
Db 1846 CAAATATCTTCAATTCGAGTACCTGCTGATTTATTCATATTCGCCGTTTAAATGCATGATG 1905
Qy 309 -----AlaAlaLeuAsnAlaGlyLys----- 315
Db 1906 AACCACTTCTATATGCTCAGAAATCGCGCTAAAACTCGCGCTTACTCAACAGAGATGCT 1965
Qy 315 ----- 315
Db 1966 TGGAGTACAGCAGCAAAAGTATGCAGATAGTACAGAGTACAGGTTAAATTAAGCTTCAAAAT 2025
Qy 315 ----- 315
Db 2026 AACTTCACATGAATTCGCTATTCAGCTTCCTGATGCTTTTATTAAGCAGCAGGATTAATGAA 2085
Qy 316 -----ArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal--- 332
Db 2086 TTTCACCAAGCCGCGACTTCTTTCGTTAAACGCTTCTACAGAGACTAATCAATCAGTTGGG 2145
Qy 333 -----ArgSerLeuValSerAlaThrSer-----LeuThrLysAsnGly 345
Db 2146 TTTCCTACCTCCCTTTTACTCTGCAGTAACACACTGCACACCTTTTGGTCAAGTAAGATACG 2205
Qy 346 LeuAlaLeuAlaGly-----GlyTyrAlaGlyValSerLysLeuGlnLysMet 361
Db 2206 CCGACATTTAGCCAGCAGCTTAATTTTGGCACACAGCTGCATGGGCGCA---GAACTTTTA 2262
Qy 362 AlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGln----- 377
Db 2263 GCGACAGCTGGGCTAGAACCTGATAGACTCGTGCAGCAGTCCCTCAAGAGTTCTTAAA 2322
Qy 378 -----LeuSerAsnLeuValGlySerValGlyValPheAlaGly 390
```

```
Db 2323 ACTGCTCTCGAGGGTTCGTTGTTTCACACTTAACTTGGCGGTCATTTGCTGTTGGG 2382
Qy 391 TrpThrThrAlaGlyLeuAlaThrAspProAla---ValLysLysAla----- 405
Db 2383 TGGGCGCTAGAAATCAATTAATTTCTTCTTTTGCAGAAAGCTTAAAAAGCTAAAGATGATTT 2442
Qy 406 -----GluSer 407
Db 2443 GAGCAGACCCAGCAAAACCAATGTCGAGCAAAATTCAGCACAATTAAGACTCCACTGTAAA 2502
Qy 408 PheIleGln-----AspLysValLysSerThrAlaSerSerThrThrSer 422
Db 2503 CTAAATACACCAATATTAAGAGCTTCAAAAAGTTAAAGAGTCAAGATCTTTAACTTCAGAT 2562
Qy 423 TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val 435
Db 2563 GAAGACCAAGAAATATCTTCAGTCACTCAGCAATTCGACAAACCTTCCCTGCATTAAGTT 2622
Qy 436 Lys-----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr 453
Db 2623 AAAGGCTATGATTTCTCAAGCAAAATGCATTTTAAGACAAATTAAGAGCTTGA AAAAGCG 2682
Qy 454 ValAsnAsnLeuArg 458
Db 2683 ATTGAGAAATACATAA 2697
```

RESULT 7

```
US-09-738-626-2739
; Sequence 2739, Application US/09738626
; Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAMA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHITO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2739
LENGTH: 8979
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2739
```

Alignment Scores:

Pred. No.:	0.000469	Length:	8979
Score:	138.00	Matches:	119
Percent Similarity:	37.168	Conservative:	85
Best Local Similarity:	21.688	Mismatches:	219
Query Match:	5.738	Indels:	126
DB:	9	Gaps:	23

US-09-825-414-66 (1-487) x US-09-738-626-2739 (1-8979)

```
Qy 7 AlaGlnGlnProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 3073 GCACACACCGCACCTGACGTGCTGGCCCGCGCATGCGCACAGCAAGTTTCGCTCCGTA 3132
```

OY	27	AlaSerSerSerValAlaGlySerValSerThrThrSerCysArgAspLeuGlnAlaLeuThr	46
Db	3133	AAGTCGGAGATCAATCCAGGACCGATTCGCGATTCGGTGAAGGACATGCTGTCCCTG	31929
OY	47	AspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAsp	66
Db	3193	GTTCACTGGAGGACCCATTT-----GTGCTCAAGTCCGATGCT	3231
OY	67	GluArgAspAlaAlaLeuAlaAlaHisAsnGluGlnIleAspAlaLeuValGluThrArgAla	86
Db	3232	CCAAACCGACGGCGCGGTGAAGGATTTCCGCGACATGCCGATGAGGTGATGCATCCGACCTG	3291
OY	87	AsnArgLeuTyr-----SerGluGlyGluThrProAlaThr	98
Db	3292	GGTCGCGCTTCGGATCCGCGCGAGAAATCCCGGACGAGAACGACCACTGATTCCTTCC	3351
OY	99	IleAlaGluThrPheAla---LysAlaGluLysPheAspArgLeuAlaThrThrLysAsp	117
Db	3352	TTGGCTGACGCTTTCCGATCCCGGAGCCGAAAGGACAGCTGTGCGACGCCAACACCC	3411
OY	118	SerAlaPheGlu-----AsnThrProPheAlaAlaLysSerValLeuGlnTyrMet	134
Db	3412	TCCGACATGCCAACCCACCGCGAGAACCCGACGCTGACGTCCGGGACATGCCAAGCTGTT	3471
OY	135	GlnPro-----AlaIleAsnLysGlyAspTyrLeuAlaThrProLeu	148
Db	3472	GCACCTGAATCCATCCGCGCATTCGCTGATTCGCGTAC-----	3513
OY	149	LysProLeuThrProLeu-----IleSerGlyAlaLeuSerGlyAlaMetAsp	164
Db	3514	-----GGCAACCCAAATTCACGTCCTGATGTGGCGCTTCGCTGCTGTGTCGCAAGT	3567
OY	165	GlnValGluThrLysMetLeuAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer	184
Db	3568	GTGATCCGAGCACCGGATGCGACCTGCGCATCGCGATGCGAGTGAATGATGCGCGGTACACATTC	3627
OY	185	ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis-----	199
Db	3628	AACGATGGACGATCCAACTCCCGGAGGCCAAAGCTGTGAATACACCGCAACCATGCTG	3687
OY	200	SerProAlaLeuGluArgGlnValValAspMetGlyLe-----AlaValIleThr	216
Db	3688	GCACCACTCTCCAGGTGAAGAAATTTGAATTCAGCTTGACAGCTCCGCACTGGACAAAC	3747
OY	217	PheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSer	236
Db	3748	CGCCAGCAATGGAGAGAGTCCGCGACCGTATCCGCA-----AC	3786
OY	237	ValGlnGluAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla-----Asn	254
Db	3787	GTCAACGGCAACTTA-----GTGCTTACCGCACCGCTGTGTGGCACTCCCACT	3837
OY	255	AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGluIleuArgGlyGlyAla	274
Db	3838	ACTTTCACGCGATCCGACGGCCAGGGCATTCAGTCCCAAGGC-----	3879
OY	275	PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr	294
Db	3880	-----ATGGGTATGGAAGACCGCGCTAAGTCCAGCAGCTCGCGGTATCTGGGACCGC	3933
OY	295	LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly	314
Db	3934	GCCGATCCACACACCCGCAATAAG-----CTGCGCTTCTCCATCTGCT	3975
OY	315	LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaAlaArg---	333
Db	3976	GAATTCGGGAAAAACAACCCACCGGAAGTAAACCGTGGCAGGAGAAAGTTCTTCCACCCA	4035
OY	334	---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaGlyGlyTyr	352
Db	4036	GAGCGCGTTTGTATCCACACCGATTCACCCAGAGTGGGATGGCAACTGCT---GGCGTT	4092
OY	353	AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLys	372

Db	4093	
		GCCTGAGATCGCGTAATGCGTGAACACATGCGCTTGAACGACGCTGCAATCTTGGCTGGA	4152
Qy	373	AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThr	392
		: : : :	
Db	4153	CACCTCCGTTGGTGAATACACCGCGCTTGTGATATGCTGGTGGTCTCCCTGGAAATCC	4212
Qy	393	Thr-----	406
		AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu	
Db	4213	GTTCGTGAGATCGTTTACCGTGTGGCTTGACCATGACACCGGCTTGATCGGATCGGATGAA	4272
Qy	407	SerPheIleGln-----	417
		AspLysValLysSerThrAla	
Db	4273	AAGGCTGTGTCACACTACGCGCTCGACGCTTTCGCCCCACACAGATGGTGTGACCCCA	4332
		: : : :	
Qy	418	SerSerThrThrSerThrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp	437
		: : : :	
Db	4333	GACAAACGTTTCGATTAGCTTGGC-----	4368
		TCGTGTTCCGAA	
Qy	438	MetSerGlyGlu-----	443
		AlaIle	
Db	4369	GCTTCGCGGTAATTCCTGAGATCGTTAACTACACACTTGGCTGGCTGCACAGACAGT	4428
Qy	444	SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuAlaGlnHisArgSerAlaPro	463
		: : : :	
Db	4428	GCTGACACCCAGCGCTGCTT-----	4458
		GCGGCCCTT	
Qy	464	GluAlaAspIleGlnGlu-----	478
		: : : :	
Db	4459	CGTGCCGATGTTGACAGACCGGTCACACAGTCAAGCTGCTTCATTTTATGATCCCTGGCAT	4518
Qy	479	GluThrProPheGlnLeuArgArgLeu	487
		: : : :	
Db	4519	GACGTGCATTCACACTCCTCCACAGCTG	4545
RESULT 8			
US-09-379-931-6			
Sequende 6, Application US/09379931			
Patent No. US20020009792A1			
GENERAL INFORMATION:			
APPLICANT: Smit, John			
APPLICANT: Bingle, Wade H.			
APPLICANT: No. US20020009792A1a1a1a1, John F.			
TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CA			
FILE REFERENCE: 08106/002003			
CURRENT APPLICATION NUMBER: US/09/379, 931			
CURRENT FILING DATE: 1999-08-24			
PRIOR APPLICATION NUMBER: US 08/614,377			
PRIOR FILING DATE: 1996-03-12			
PRIOR APPLICATION NUMBER: US 08/194,290			
PRIOR FILING DATE: 1994-02-09			
PRIOR APPLICATION NUMBER: US 07/895,367			
PRIOR FILING DATE: 1992-06-09			
NUMBER OF SEQ ID NOS: 12			
SOFTWARE: FASTSEQ for Windows Version 4.0			
SEQ ID NO 6			
LENGTH: 3300			
TYPE: DNA			
ORGANISM: Caulobacter crescentus			
FEATURE:			
NAME/KEY: CDS			
LOCATION: (101)...(3178)			
US-09-379-931-6			
Alignment Scores:			
Pred. No.: 0.00475			
Score: 132.00			
Percent Similarity: 36.51%			
Best Local Similarity: 20.95%			
Query Match: 5.48%			
DB: 10			
Gaps: 23			

US-09-825-414-66 (1-487) x US-09-379-931-6 (1-3300)

Query Match: 5.42% Indels: 105
DB: 10 Gaps: 17
US-09-825-414-66 (1-487) x US-09-815-242-4580 (1-7104)

QY 4 AsnGlnSerAlaGlnInProProGlyValAlaMetGlnSerPheArgThrAlaSer--- 22
DB 2104 AATCATCATATACAAAGATGAAAGAAATGTAATAATGTAATAATGTAATAA 2163
QY 23 AsnAlaSerLeuAlaSerSerSerValArgSerValSerThr 36
DB 2164 AAGCAATTAAGATATTGATGACAGCAACAATGCAAGATGCAATTAATAA 2223
QY 37 ThrSerCysArgPheLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAla 56
DB 2224 AAACGATCATGATATTATATCAAACTGCACCTGCTACAAACAGCAAAAGCAGCTCTT 2283
QY 57 HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala 73
DB 2284 GAAGATTTGACGAGATGTTTCAAGCACAAATGATGATCAAGCCTTTAAATCTGATACA 2343
QY 74 HisAsnGlnGlnIleAspAlaLeuValGlnThrArgAlaAsnArgLeuTyrSerGluLys 93
DB 2344 ACAATGACAGAGTACGCGAAGCTATGAA-----CGATTAAATGACGCTAAA 2391
QY 94 GluThrProAlaThrIleAlaGlnThrPheAlaLysAlaGlnLysPheAspArgLeuAla 113
DB 2392 GTTCTGCTGTTAAAGCAATGTAAGCGCACAGACTGCACAAAGATTTGAAGAAGTTAAA 2451
QY 114 ThrThrAlaSerSerAlaPheGlnAsnThrProPheAlaAlaAlaSerValLeuGlnTyr 133
DB 2452 AACGAGAAATCTCAAAAAATTTGAAAT-----ATTACTGACTCTACTCAACA 2499
QY 134 MetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrPro 153
DB 2500 AAATGATGCTGCTAATGAAGATTAACACAGCAGCAACGCTAGAAAATCCAAATCT 2559
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg 173
DB 2560 ACAGTTTCAAAATGCAACTAAT-----GAGAGATG----- 2589
QY 174 AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla 193
DB 2590 GCAGAGCTGATGACAGCAGTGAAGACCTCAAAAGCAAGTTTACAGATCCAACTT 2649
QY 194 ValSerValLysArgHisSerProAlaLeuGlnLysGlnValAlaLysMetGlyIleAla 213
DB 2650 GTTAATGCAAAAGAGAAAGTGTGATACAAATCAAAAGATTAAGTAAATCAATGCA 2709
QY 214 ValGlnThrPheSerAlaLeuAsnValAlaArgThrValLeuAlaProAlaLeuAlaSer 233
DB 2710 ATTCAAAA----- 2718
QY 234 ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyLysValAla 253
DB 2719 CAAAGCAAAAGTTAAACCTGCGATACGAGAGAGAGAAAGCGA----- 2763
QY 254 AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyLys 273
DB 2764 -----TATATATACACGTTAAACAGAAATTCAAATATACAAAT----- 2799
QY 274 AlaPheValLeuGlyMetLysAspLysGlnProLysAlaAlaLeuSerGlnGlnThrAsp 293
DB 2800 -----GCTTCAACTACAGAGAAACAAAGCTGCATATACAGAA----- 2838
QY 294 TrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
DB 2839 ---TTAGATTACTAAAGACAGAAAGCAAGAACTTTGATGCTGCAAAATACAAACAGT 2895
QY 314 GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys 330
DB 2896 -----GATGTAACACAGCTAAAGACAAATGATATTGCT 2928

QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
DB 2929 GCATTTAATCAAGTACACGCGCAACACTAAGAAATCGATGCTAAAGCGGAATGCT 2988
QY 351 GlyTyrAlaGlyValSerLysLeu-----Gln 359
DB 2989 CAAAAGCAAGTGAACGTAACAACTGCAATTGACAAATGAAATGATGCTACTGAAAGAA 3048
QY 360 LysMetAlaThrLysAsnIleThrAspSerValThrLysAlaAlaValSerGlnLeuSer 379
DB 3049 CAAACAGCTGCAGAAAGATAAAGCTTGAATCAGCACTAGTACTGCAAAACGCTGATATGAT 3108
QY 380 AsnLeuValGlySerValGlyValPheAlaGlyTyrPThrThr-----Ala 394
DB 3109 AATGCTGACGCAAAATFACGATGATGATTAATGCAAAACTACTATATGAAGCTACAAATGCA 3168
QY 395 GlyLeuAlaThrAspProAlaValLys---LysAlaGlnSerPheIleGlnAspLysVal 413
DB 3169 GCCATTACCTGATGCAAAATGTTAAACCAACAGCGAAACAAAGCAAAATGCTGATTAAGTA 3228
QY 414 LysSerThr-----AlaSerSerThrThrSerTyrVal 424
DB 3229 CAACGCGAAGAAACGCAATGATGCTAATATACGCTGCACACACAGAAAGAAAGCAGCT 3288
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlnAlaIleSer 444
DB 3289 GCGAAACACAAAGTTTCAACGAAAGAAACAAACAGCTGATACGAAATGATGATG---GCA 3345
QY 445 SerThrGlyAlaSerLeuArgSerThrValAsn 455
DB 3346 CATACAAATGCAAGAGTTGAAGCGCTTAATAAT 3378

RESULT 10
US-09-815-242-8291
Sequence 8291, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8291
LENGTH: 7107
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(7107)

US-09-815-242-8291

Alignment Scores:

Pred. No.:	0.00201	Length:	7107
Score:	130.50	Matches:	99
Percent Similarity:	33.208	Conservative:	64
Best Local Similarity:	20.16%	Mismatches:	223
Query Match:	5.42%	Indels:	105
DB:	10	Gaps:	17

US-09-825-414-66 (1-487) x US-09-815-242-8291 (1-7107)

```
QY 4 AsnGlnSerAlaGlnGlnProGlyValAlaMetGluSerPheArgThrAlaSer----- 22
DB 2104 AATGATCATTTACAGATGAAAAAGATGACCAATGATTAATTTGTAATTTGAACA 2163
QY 23 -----AspAlaSerLeuAlaSerSerSerValArgSerValSerThr 36
DB 2164 AAGCAATTAAAGATATTTATTCAGCACACACAAATGACAAAGTAGACCACTTAATAACA 2223
QY 37 ThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaIle 56
DB 2224 AAAGCATCAATGATATTTATTAATCAACTGCTACCAAGCTTAAGCAGCAGCTCTT 2283
QY 57 HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala 73
DB 2284 GAAGATTTTGACGAGTTGTTCAAGCACAATTTGATCAACGCTTTAATCCTGATACA 2343
QY 74 HisAsnGlnGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGly 93
DB 2344 ACAAAATGAAGAGTAGCCGAGACGATTTGAA-----CGTATTAATGACAGCTAAA 2391
QY 94 GluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla 113
DB 2392 GTTTCGTCGTTTAAACCAATTGAGGACACAGACTGCACAGATTTTAAAGAGATTTAA 2451
QY 114 ThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr 133
DB 2452 AACGAAGAATGCTCAAAAATTGAAAT-----ATTACTACCTCTACTCAACA 2499
QY 134 MetGlnProAlaIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrPro 153
DB 2500 AAAATGATGCTATATATCAAGTTAAACAAGCACAAGCTAGAAAACCTCAAAATGCT 2559
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetLeuAspArg 173
DB 2560 ACAGTTTCAAAATGCAACTAT-----GAAGAAGTA----- 2589
QY 174 AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla 193
DB 2590 GCAGAGCTGATGACGACATAGACAGCAGCTCAAAAGCAGTTTACATGACATCCAAATT 2649
QY 194 ValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaValAspMetGlyIleAla 213
DB 2650 GTTAATCAAAACAGCAAGCTGCTGATACAAATCAAAAGTATTACATTAATCAATGCA 2709
QY 214 ValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSer 233
DB 2710 ATTCAACA----- 2718
QY 234 ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyLeuValAla 253
DB 2719 CAAGCAAAAGTTAAACCTGACCTGATACGGAAGTAGAAAACGCA----- 2763
QY 254 AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly 273
DB 2764 -----TATATATACAGCTAAACAAGAAATTCAAAATGCAAT----- 2799
QY 274 AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp 293
DB 2800 -----GCTTCAACTACAGAGAAAACAAGCTGCATATACAGAA----- 2838
QY 294 TrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
```

```
DB 2839 ---TTAGATACTTAAAGAACAGACAGACAGAACAAATCTGATGCTCAAAATACAAACAGT 2895
QY 314 GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys 330
DB 2896 -----GATGTAAACAACAGCTTAAACAGACATGATTTGCT 2928
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
DB 2929 GCATTTATCAAGTACACAGCGGCACACTAAGAAATCGGATGCTTAAACGGAAATCGCT 2988
QY 351 GlyTyrAlaGlyValSerLysLeu-----Gln 359
DB 2989 CAAAAGCAAGCAGCACTTAAACTGCATTTGAAGCAATGATTCGACTACTGAAGAA 3048
QY 360 LysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSer 379
DB 3049 CAACAAGCTGCAAAAGATTAAGTTGATCAAGCAGTAGTACTTCAAAACGCTGATATAGAT 3108
QY 380 AsnLeuValGlySerValGlyValPheAlaGlyTrpThr-----Ala 394
DB 3109 AATGCTGACGACATATCTGATGATGATTAATGCAAAAACCTACTATGACACTACATCGCA 3168
QY 395 GlyLeuAlaThrAspProAlaValLys---LysAlaGluSerPheIleGlnAspLysVal 413
DB 3169 GCATTAACAACCTGATGCAATGTTAAACCAACAGCAAGCAATTCGCTGATTAAGTA 3228
QY 414 LysSerThr-----AlaSerSerThrThrSerTyrVal 424
DB 3229 CAAGCCCAAGAAACAGCAATTCATGCTTAATTAACGCTGCAACAACAGAAAAAGCAGCT 3288
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSer 444
DB 3289 GCGAAACCAACAGTTCAAACTGAAAACACACAGCTGATACGACATTTGCTGCT---GCA 3345
QY 445 SerThrGlyAlaSerLeuArgSerThrValAsn 455
DB 3346 CATACAAATGACGAAGTTGAGCGGCTAAAAAT 3378

RESULT 11
US-09-815-242-4760
; Sequence 4760, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 4760
LENGTH: 6228
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-815-242-4760

Alignment Scores:
Pred. No.: 0.00212 Length: 6228
Score: 129.50 Matches: 106
Percent Similarity: 35.88% Conservative: 91
Best Local Similarity: 19.31% Mismatches: 201
Query Match: 5.38% Indels: 151
Gaps: 22

US-09-825-414-66 (1-487) x US-09-815-242-4760 (1-6228)

Oy 20 ThrAlaSerAspAlaSerLeu-----AlaSerSerSerValArgSerValSer 35
Db 2788 ACTGCAGCAAAACAAACATTTTAAAGCTAATACCAATGCTGATGTGAACAAAGTAAAG 2847
Oy 36 ThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAla 55
Db 2848 ACAATGCGATTCAAGGATACAAACATTTACACAGCTACAA-- 2892
Oy 56 AlaHisArgPheSerValIleGlySerProAspGluArgAspAlaLeuAlaHisAsn 75
Db 2893 -----GTAATAACAGATGCATAAAATGCC----- 2916
Oy 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyLuhTr 95
Db 2917 -----ATCGATAAAAGTCGGAAAGCAACATAATACGATATTTAATATATATAT 2967
Oy 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
Db 2968 ---GGCAGCTCGAAGAACAAACACACACAACTACTGATCAAGCTGACCA 3024
Oy 116 AlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGln 135
Db 3025 GCGAAGCAAAATATTAAT-----GCAGCAGATACCAATCAAGAAAGTTGCA 3069
Oy 136 ProAlaIleAsnLysGlyAspTyrLeuAlaThrProLeuLysProLeuThrProLeuIle 155
Db 3070 CAAGCAAAAGATCAGGACACACAAATATAGTAGATTCAACCGGCAACA----- 3120
Oy 156 SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-----MetMet 171
Db 3121 -----CAAGTTAAACGGATACGCGCATGTTTAAAT 3153
Oy 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 3154 GATTAAGCGCGAGAGCGCATTAACAATATCAATGCTACAACTGGCGCAGCTCGAAGAG 3213
Oy 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGly 211
Db 3214 AAACAGAGAGCATTAATCGTCTCACTTAAATAATAGAGCATTAATGATATGCT 3273
Oy 212 IleAlaValGlnThrPheSerAlaLeuAsnValValArgThr----- 225
Db 3274 GTG---AGCTCTACTACTGCGGTGCTCAATAGTATTAGACAGCATGCAATCAATC 3330
Oy 226 ---ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla----- 240
Db 3331 GGGCAGATTCAACCGCATGTAAAGAAACAAACTGCTACAGGTGTATTAAATGATTTA 3390
Oy 240 ----- 240
Db 3391 GCAACTGCTAAAAACAGAAATTAATCAAAACAAATGCAACAACTGAAGAAAGCAA 3450
Oy 241 -----ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAla 255
Db 3451 GTGGCTTTAAATCAAGTGAGTACAGAGTTAGCAAGGCA-----ATTATATATATA 3501
Oy 256 GlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPhe 275

Db 3502 AATCAAGCTGATACAAATCGCGAAGTAGATCAAGCGCAACATTAAGTACAAAACATTT 3561
Oy 276 -----ValLeuGlyMetLysAspLysGlu 283
Db 3562 AATGCGATTCAAGCCAAATATTGTTAAAAACCTGACGATTAAGCAACAATCAATCAGCAT 3621
Oy 284 ProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaTyrLysAlaIleLysSer 303
Db 3622 TATATGCTAAATTAATTCGGAATATCAATGCTACACAGATGCAAGAAATGATGACAAAAT 3681
Oy 304 AlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla----- 318
Db 3682 GCTGCGATCAAT-----ACTTTAATCAAGACAGACAAACAACTATTGAAGATTAA 3735
Oy 319 -----GlyLeuProLeuAsp-----ValAlaThrAspGlyLeuLys 330
Db 3736 CAAGTAAACAAATGCAAGATGACCAAGCTGCGACAGTAGACAGAAATTAATTCAT 3795
Oy 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
Db 3796 GGTGTCAAGTTGATGTAGTAAAAAACACAGCGCGCATTAATCAATCAATCAATCAAT 3846
Oy 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAla-----ThrLysAsnIleThrAsp 368
Db 3847 -----GCTGAAGTGGCGACAGCGATTGACCGTTAAACAAACACCTTAATGCAACTGC 3900
Oy 369 SerAlaThrLysAlaAlaValSerLysLeuValGlySerValGlyValPhe 388
Db 3901 GACAAAAGCAGGCTGCTGTTAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 3939
Oy 389 AlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAla---GluSer 407
Db 3940 -----AAAGATCAAGCAATTAATCAATTAATCAATTAATCAATTAATCAATTAAT 3972
Oy 408 PheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal----- 424
Db 3973 CAACAAATGATCAAGTGTGACACCACTCAATCAACAGCGTAAATGCTATGATTAATGTT 4032
Oy 425 ---AlaAspGlnThrValLys-----LeuAlaLysThrValLysAsp 437
Db 4033 GAAGCTGAAGTAGTATTAATTAACCAAGCAATTCGACATATTGAAAAAGCTGTTAAAGAA 4092
Oy 438 MetSerGlyAlaAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeu 457
Db 4093 AAGCAACAGCAAAATGATTAAT-----AGCTTGATTTCAACAGATTAATGAGAAA 4140
Oy 458 ArgHisArgSerAlaProGlnAlaAspIleGluGlyGlyIleSerAlaPheSerArg 477
Db 4141 GAAGTTGCTTCAACAGATTGCTTAAGAAAAAGAAAAAGCACTGACGATTGACCAA 4200
Oy 478 SerGluThrProPheGlnLeuArgArg 486
Db 4201 GCTCAACGAATAGTCAGGTGAATCAA 4227

RESULT 12
US-09-815-242-8815
Sequence 8815, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: EITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21

```
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 8815
LENGTH: 6561
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(6561)
US-09-815-242-8815

Alignment Scores:
Pred. No.: 0.00228      Length: 6561
Score: 129.50          Matches: 106
Percent Similarity: 35.88%      Conservative: 91
Best Local Similarity: 19.31%    Mismatches: 201
Query Match: 5.38%             Indels: 151
DB: 10                      Gaps: 22

US-09-825-414-66 (1-487) x US-09-815-242-8815 (1-6561)
QY 20 ThrAlaSerAspAlaSerLeu-----AlaSerSerValArgSerValSer 35
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 2788 ACTGCAGCAACACACAACTTTAAACGCTAATATACCAATGCTGTGACAAAGTAAAG 2847
QY 36 ThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuShiHisValPheAla 55
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2848 ACAATGCTCATTCACGAATACCAACAAATTAACACGCTACACAA----- 2892
QY 56 AlaHisArgPheSerValIleGlySerProAspGluArgAspAlaIleuAlaHisAsn 75
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2893 -----GTAATAACAGATGCAAAAATGCC----- 2916
QY 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGluThr 95
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2917 -----ATCGATTAAGTACGCGAAGCAACATTAATATGATATTAATATGAT--- 2967
QY 96 ProAlaIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 2968 ---GCCACGCTCGAAGACAAACACAGACACAACTTACTTGAAGCTGTAGCCACA 3024
QY 116 AlaSerSerAlaPheGluAsnThrProPheAlaAlaSerValLeuGlnTyrMetGln 135
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3025 GCGAACCACAAATATTAAAT-----GCACGAGATCGAATCAAGAGTTGCCA 3069
QY 136 ProAlaIleAsnLysGlyAspTyrPheAlaThrProLeuLysProLeuThrProLeuIle 155
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3070 CAAGCAAAAGATCAGCGCACACAAATAATATAGTATGATTCACCGGCACACA----- 3120
QY 156 SerGluAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-----MetMet 171
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3121 -----CAAGTTAAACCGAATCTGCGCAATGTTGTAAT 3153
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3154 GATTAAGCCGAGAGCGATTAACAATATATCAATGCTACAACTGGCCGAGCTCGAGAAGAG 3213
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMetGly 211
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
Db 3214 AAACAAGAACGATTAATGCTGCAATACACTTAAAAATAGACATTAATGATATTGGT 3273
QY 212 IleAlaValGlnThrPheSerAlaLeuAsnValAlaArgThr----- 225
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3274 GTG---ACGTCTACTACTGCGATGATGATCAATAGATTAGACAGCATGCAATCAATC 3330
QY 226 ---ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla----- 240
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3331 GCGCGAGTTCAACCGCATGTGACGAAGAACAAACTGCTACAGGTATTAAATGATTTA 3390
QY 240 ----- 240
Db 3391 GCACAGTCTAAAAAGCAGAAATTAAATCAAAACAAATGCAACACTGACAGAAAAAGCA 3450
QY 241 -----ValAspPheGlyValSerThrAlaGlyLeuValAlaAsnAla 255
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3451 GTGCGTTTAATCAATGCTGTCAGAGATTAGACACGCA-----ATTAATTAATATA 3501
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAlaPhe 275
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3502 AATCAAGCTGATACAAATGCGGAATGATCAAGCCGCAACATTTAGTACAAAACAAATT 3561
QY 276 -----ValLeuGlyMetLysAspLysGlu 283
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3562 AATGCGATTCAGCCAATATTGTTAAAAACCTGCAAGCATTAACACAAATCAATCAGCAT 3621
QY 284 ProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaTyrLysAlaIleLysSer 303
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3622 TATTAATGCTAAATTAATCTGTAATCAATGCTACACCGATGCAAGATGATGAAAAAT 3681
QY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla----- 318
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3682 GCTGCGATCAAT-----ACTTAATATCAAGACAGACACACACACTTGTGAAATATTAA 3735
QY 319 -----GlyLeuProLeuAsp-----ValAlaThrAspGlyLeuLys 330
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3736 CAAGCTAACCAAAATGAGAAATGACCAAGCTGCGACAGTACAGATTAATATGAT 3795
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3796 GCTGTTCAGATTGATGATGATTAACAAACAGCGCGAGATTAATATCACT----- 3846
QY 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAla-----ThrLysAsnIleThrAsp 368
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3847 -----GCTGAGTGGCGAAGCGCTATTGAAGCGGTAAACAACTAATGCAACTGAC 3900
QY 369 SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPhe 388
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3901 GAAGAAAGAGCGAGCTGCTGTTATCAAAATCAATCAACTT----- 3939
QY 389 AlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAla---GluSer 407
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3940 -----AAGATCAACGATTAATCAATTAATCAAAAC 3972
QY 408 PheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal----- 424
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3973 CAACAACAAATGATAGTAGACACACACTACAAATCAACGCGTAAATCTATAGTAAATGTT 4032
QY 425 ---AlaAspGlnThrValLys-----LeuAlaLysThrValLysAsp 437
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 4033 GAAGCTGAGTAGTAATTAAACCAAGCAATTCAGATTAATGAAGAAAGCTGTTAAAGAA 4092
QY 438 MetSerGlyGlnAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeu 457
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 4093 AAGCAACAGCAAAATTTGATAT-----AGTCTGATTCACACAGTAATATGAGAAA 4140
QY 458 ArgHisArgSerAlaProGluAlaAspIleGluGluGlyGlyLysSerAlaPheSerArg 477
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 4141 GAAGTGTCTTCAACACATTAAGTAAAGAAAAAGAAAAAGAAAGCACTTCAGCTATTGACCAA 4200
QY 478 SerGluThrProPheGlnLeuArgArg 486
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 4201 GCTCAACGAAATGTCAGTGAATCAA 4227
```

```
RESULT 13
US-09-815-242-8615
; Sequence 8615, Application US/09815242
; Patent No. US20020061569A1
GENERAL INFORMATION:
; APPLICANT: Haseibeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8615
LENGTH: 7035
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(7035)
US-09-815-242-8615

Alignment Scores:
Pred. No.: 0.00283 Length: 7035
Score: 129.00 Matches: 96
Percent Similarity: 38.528 Conservative: 97
Best Local Similarity: 19.168 Mismatches: 231
Query Match: 5.36% Indels: 79
DB: 10 Gaps: 12

US-09-825-414-66 (1-487) x US-09-815-242-8615 (1-7035)
QY 4 AsngInSerAlngInlProProGlyValAlaMetGlSerPheArGThrAlaSerAsp 23
DB 4204 TCMAACTCTGCGACGAAAGTGCCTCAACCTCAACAGTACAGCGAATCCGAT 4263
QY 24 AlasSerLeuAlaSer-----SerSerValArGSerValSerThrPrhrSer 38
DB 4264 TCACAAAGCAGCATCATATATCAAGTCAAGCAGCAAGCGAATCCGATCAAC 4323
QY 39 CysArGAsPleuGlAlaIleThrAsp-TyrlEuLys-----HisHisValPheAl 55
DB 4324 TCACGCTACTAGGATTCACAGATATATCTAAAGTACAAAGCCAAATCAGATTCA 4383
QY 55 AlaIleAsrGpHeSerValIleGlySerProAspGluArGAsPAlaAlaLeuAlaHisAs 75
DB 4384 ACCACATCT-GCATCATTTAAGGTTCAGAAAGCGAATCCGATTCACAAAGCATATCAAC 4442
QY 75 nGluGInIleAsPAlaLeuValIGluThrArGAlaAsnArGLeuTyrlSerGluGlyGluTh 95
DB 4443 AAGCACAGTGAAGTCAAAATCATCAAGATACATCCATCCAGTCAAGTCAACAGTAC 4502
```

```
QY 95 r-----ProAlaThrIleAlaGluThrPheAlaLysAlaGlu 108
DB 4503 AAGCAACTCAGGATCAGCAAGTACGTCACATTCGCTCACTACATCAGCAAGGCAAGTA 4562
QY 108 sPheAsPArGLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaI 128
DB 4563 ATCCGATTTCGCTGTCACATCATTAAGTATTCAGATTCAGCGTCAATGCAGCAAGTGA 4622
QY 128 aSerValLeuGInTyrlMetGlnProAlaIleAsnLysGlyAspTrPleuAlaThrProLe 148
DB 4623 ATCCGATTTCACAAAGCAGCATCAACGCTAAGTATTCACAAAGTACATCAATCAAT 4682
QY 148 uLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValIGlyTh 168
DB 4683 TCGA---ATGTCGACATTCGACAGTACGATATTCGGAATATCATGATTCGATTC 4739
QY 168 rLysMetLeAsPArGAlaArGlyAspLeuHisTyrlEuSerThrSerProAsp-LysL 188
DB 4740 AACATCAGAAAGTACAAAGTATCCGATTCGACATCAATCATTAAGCATTCACAAAG 4799
QY 188 euHisAsPAlaMetAlaValSerValLysArg----- 198
DB 4800 CACATCAGAAAGTACAAAGTATCCGATTCGACATCAATCAATCAATCAATCAATCAAT 4859
QY 199 -----HisSerProAlaLeuGlyArGAlaValAlaAspMetGlyIleAlaValGlnT 216
DB 4860 TCGTAGTACATCAGCTTCACATCAGTACCTGATTCGATTCAGTACATCAGTTCATCA-CAA 4918
QY 216 hPheSerAlaLeuAsnValValArGThrValLeuAlaProAlaLeuAlaSerArGProS 236
DB 4919 GTATGCGCTTTCAACAGTACATCAACAGTACATCAACAGTACATCAATTCATTCGATA 4978
QY 236 eValIGluGlyAlaVal-----AspPheGlyValSerThrAlaGlyGlyLeuValAla 254
DB 4979 GTGTGAGTGATTCACATCAGCTCAGCTCAGTACATCAATCAATCAATCAATCAATCAAT 5038
QY 254 snAlaGlyPheGlyAspArGMetLeuSerValIGlnSerArGAspGlnLeuArGlyGlyA 274
DB 5039 CTATATCGTTAAGTATTCGACAGCAGTACATCAATCGCTGATCAAGTAATG----- 5091
QY 274 lArPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsPT 294
DB 5092 -----AGCGCAAGCATATCTGATTCACAAAGTA 5119
QY 294 rPleuAsPAlaTyrlLysAlaIleLysSerAlaSerTyrlSerGlyAlaAlaLeuAsnAlaG 314
DB 5120 TGTCAAGATCTGTAATGATTCAGAAAGTGAAGTGAATCTAATTCGAAAGTCACTCA- 5178
QY 314 LylsArGMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArGs 334
DB 5179 --AAATCGATGAGTGT--TCACAAAGCGTAAAGTATTCGCG-----T 5218
QY 334 eLrLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyAlaG 354
DB 5219 CATTTAGCGTCTTCAACGCTCATTAAGAAAA----- 5247
QY 354 lYValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAla 374
DB 5248 -----TCAGAAAGGTAAAGCGAGTCTTTTCATTTGAGTGG 5281
QY 374 lAvalSerGlnLeuSerAsnLeuValIGlySerValIGlyValPheAlaGlyTrpThrThra 394
DB 5282 GGTCAACATCGATGATGATTCAGTAACACACAAG-----GATTCGT 5323
QY 394 lAglyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValL 414
DB 5324 CATCATTTAAGTATTCGACGCTCACAAGAGTGAAGAGGTGAGTGAATCTGATTCAT 5383
QY 414 ySerThrAlaSerSerThrThrSerTyrlValAlaAspGlnThrValLysLeuAlaLysT 434
DB 5384 TAAAGTATTCAAATCAACAGTGTGCTTCATCAACAGTACATCAGTTCATTCATTCAGCA 5443
```


QY 414 LysSer----- 415
Db 3241 CAAGCTCAGAAACAGCAATGATGAAATACGCTCAACACTGAAGAAAAGACACT 3300
QY 416 -----ThAlaSerSerThrThrSerTyrValAlaAsp----- 426
Db 3301 GCTAACACACAGTTCAAACCTGAAAAACACACACTGATGCCGCAATGACGACACAT 3360
QY 427 -----GlnThrValLys 430
Db 3361 ACAATGCGAAGTTGAAGCGCTAAAAAGACCAATGCTTAAATTTGAAGCGATTGAC 3420
QY 431 LeuAlaIsthrValLysAspMetSerGlyAlaIleSerSerThrGlyAlaSerLeu 450
Db 3421 CCAGCAGACACACACTAAAGATATGCGAAGACCAATTTGCTAGAAAAGCAATGAACGT 3480
QY 451 ArgSerThrValAlaAsnIleuArgHisArgSerAlaProGlu-----Ala 465
Db 3481 AAACAGCAATCGCTCAAAACCGCAAGACATTTACTGTAAGAAATTCGACGCGTAATGCG 3540
QY 466 AspIleGluGluGlyIleSerAlaPheSerAlaPheSerGlu 479
Db 3541 GACGTAGTATGCTGCTGACACACCAATATGCAACATTGAA 3582

RESULT 15
US-09-815-242-8869
: Sequence 8869, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 8869
: LENGTH: 7437
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(7437)
US-09-815-242-8869

Alignment Scores:
Pred. No.: 0.00438
Score: 127.50
Percent Similarity: 31.77%
Best Local Similarity: 19.13%
Query Match: 5.30%
Length: 7437
Matches: 106
Conservative: 70
Mismatch: 235
Indels: 143

DB: 10 Gaps: 18
US-09-825-414-66 (1-487) x US-09-815-242-8869 (1-7437)
QY 4 AsnGlnSerAlaGlnGlnProProGlyValAlaMetGlnSerPheArgThrAlaSer--- 22
Db 2116 AATGCATCATTTACAAATGAAAAAGATGTAGCAATATTAATTTGCTAAATTTGAACA 2175
QY 23 -----AspAlaSerLeuAlaSerSerValArgSerValSerThr 36
Db 2176 AAGCAATTTAAAGATATTGATGACGACCAACAAATGACACAGTAAAGCAATTAACA 2235
QY 37 ThrSerGlyArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAla 56
Db 2236 AAACCAATCAATGATATTTAATCAAACTACACCTGCTACAAACAGCTAAAGCAGCTCTT 2295
QY 57 HisArgPheSerValIleGlySerProAspGlnArgAspAlaAlaLeu-----Ala 73
Db 2296 GAAGAAATTTGACGAAGTGTTCAGACACAAATTTGATCAAGCACCTTAAATCCTGATACA 2355
QY 74 HisAsnGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGly 93
Db 2356 ACAATGACAGAGTAGCGGAGAGCTATTGAA-----GCTATTATTCAGCTAA 2403
QY 94 GluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla 113
Db 2404 GTTCTGCTGTTTAAAGCAATTTGAAGCGACCAACGACTGACACAGATTTTAAAGAGTTAAA 2463
QY 114 ThrThrAlaSerSerAlaPheGlnAsnThrProPheAlaAlaAlaSerValLeuGlnTyr 133
Db 2464 AACGAAGAAATCTCAAAAATTTGAAAT-----ATTACTGACTTACGCCAACA 2511
QY 134 MetGlnProAlaIleAsnLysGlyAspTyrLeuAlaThrProLeuLysProLeuThrPro 153
Db 2512 AAATGATGCTGCTTATTAAGATTAAGATTAACAGCTCAACAGCTACAAAAGCTCAAAATGCT 2571
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg 173
Db 2572 ACAGTTTCAAAATGCACAAAT-----GAAGAACTA----- 2601
QY 174 AlaArgGlyAspLeuHisLysThrLeuSerThrSerProAspLysLeuHisAspAlaMetAla 193
Db 2602 GCAGAGCTGATGCGACAGCATGATGACCTCAAAAGCAAGCTTACATGACATCCAAAGTT 2661
QY 194 ValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaValAspMetGlyIleAla 213
Db 2662 GTTAATTCAAACAGCAAGTGTGATACAAATCAAAAGATATAGTAAATCAATGCA 2721
QY 214 ValGlnThrPheSerAlaLeuAsnValArgThrValLeuAlaProAlaLeuAlaSer 233
Db 2722 ATTCAACA----- 2730
QY 234 ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyLysLeuValAla 253
Db 2731 CAACCAAAAGTTAAACCTGCGTGAATACGAAATGCAAAAGCA----- 2775
QY 254 AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly 273
Db 2776 -----TATATACACCGTAACAGAAATTCATAAATGACAT----- 2811
QY 274 AlaPheValLeuGlyMetLysAspGlyGluProLysAlaAlaLeuSerGluGluThrAsp 293
Db 2812 -----GCTTCACTACAGAGAAACAAAGAAATTCATAAATGACAT----- 2850
QY 294 TrpLeuAspAlaTyrLysAlaIleLysSerAlaSerLysArgGlyAlaAlaLeuAsnAla 313
Db 2851 -----TTAGATTACTAAAGCAAGCAAGCAACAAATCTTATGCTGCAAAATACAAACAGT 2907
QY 314 GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys 330
Db 2908 -----GATGTACACACAGCTTAAAGCAATATGATTTGCT 2940
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350

